

STIC-Biotech/ChemLib

66971

From: Ton, Thaia
Sent: Monday, May 20, 2002 9:41 AM
To: STIC-Biotech/ChemLib
Cc: Ton, Thaia
Subject: Sequence Search Request

I would like to request a search for Application No. 09/696,686 [Inventor: Allen, Filing Date: 10/26/2000]

A standard sequence search for SEQ ID NOS: 45, 46, 47, 50 and 51 against the commercial nucleic acid databases.

Thank you very much.

Thaia N. Ton

Patent Examiner

Art Unit 1632

Room 12A16 CM1

Mailbox: 12E12 CM1

(703) 305-1019

RECEIVED

of Contact:
Sheppard

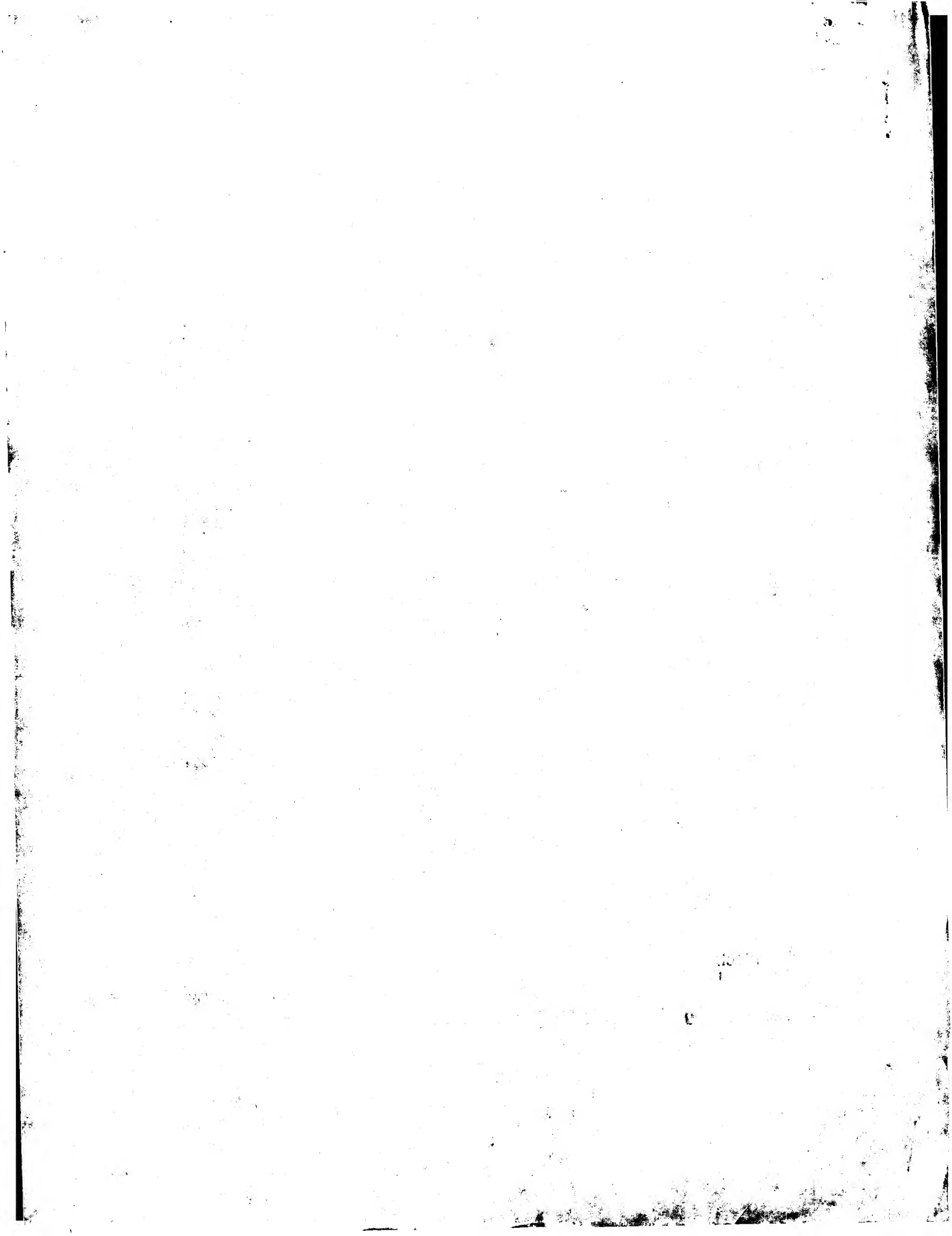
Searcher:
Phone:
Location:
Date Picked Up:
Date Completed: 5/25/02
Searcher Prep/Review:
Clerical:
Online time:

TYPE OF SEARCH:

NA Sequences:
AA Sequences:
Structures:
Bibliographic:
Litigation:
Full text:
Patent Family:
Other:

VENDOR/COST(where applic.)

STN:
DIALOG:
Questel/Orbit:
DRLink:
Lexis/Nexis:
Sequence Sys.:
WWW/Internet:
Other (specify):



Ton, Thaian

From: Ton, Thaian
Sent: Monday, May 20, 2002 9:41 AM
To: STIC-Biotech/ChemLib
Cc: Ton, Thaian
Subject: Sequence Search Request

I would like to request a search for Application No. 09/696,686 [Inventor: Allen, Filing Date: 10/26/2000]

A standard sequence search for SEQ ID NOS: 45, 46, 47, 50 and 51 against the commercial nucleic acid databases.

Thank you very much.

Thaian N. Ton

Patent Examiner

Art Unit 1632

Room 12A16 CM1

Mailbox: 12E12 CM1

(703) 305-1019



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 23:06:04 ; Search time 4203.75 seconds
(without alignments)
1187.957 Million cell updates/sec

Title: US-09-696-686-51

Perfect score: 370

Sequence: 1 tgcgattgccagcaaatgc.....atgaggacgcccctgcggcc 370

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.2	13.8	1090	12	AG128386 Pan trogl
2	50.6	13.7	941	12	AL04537 T3 end of
3	49.8	13.5	1101	12	CNS06LUR7
4	49.6	13.4	723	12	AL106944 Drosophil
5	49.4	13.4	1196	10	AZ184008 SP_1002_A
6	49.2	13.3	882	10	BF256755 HVSMEF001
7	49	13.2	660	12	AG126129 Pan trogl
8	48.6	13.1	1101	12	AZ198637 SP_1037_B
9	48.4	13.1	295	12	CNS00GUM
10	48.2	13.0	458	10	AL072271 Drosophil
11	48	13.0	843	12	BE348494 ht69l2_x
12	47.8	12.9	466	10	A0738512 HS_5381_B
13	47.8	12.9	937	10	BG520030 ps23b11.y
14	47.8	12.9	1101	12	BI416482 hasp001xa
15	47.6	12.9	922	12	AL071731 Drosophil
16	47.4	12.8	648	12	AL418302 T7 end of
17	47.4	12.8	762	12	BH021826 GH_MBB000
					AG127831 Pan trogl

18	47.4	12.8	918	12	AQ895329
19	47.4	12.8	933	12	CNS04125
20	47.4	12.8	1638	10	BG179554
21	47.2	12.8	271	12	CNS008FE
22	47.2	12.8	544	12	CNS015XA
23	47.2	12.8	780	12	B21793
24	47.2	12.8	996	12	CNS04OW3
25	47.2	12.8	999	12	CNS02ITK
26	47.2	12.8	1026	12	AG081192
27	47.2	12.8	1293	10	BF256498
28	47	12.7	810	12	AG127829
29	47	12.7	855	9	AL572700
30	47	12.7	927	12	CNS06XJF
31	47	12.7	964	12	CNS003WG
32	46.8	12.6	378	9	AL535833
33	46.8	12.6	913	12	AG076623
34	46.8	12.6	936	12	CNS01808
35	46.8	12.6	999	12	AG132063
36	46.8	12.6	1235	10	AL539262
37	46.6	12.6	511	9	AL539262
38	46.6	12.6	684	10	BG786306
39	46.6	12.6	741	12	AG049269
40	46.6	12.6	874	12	AZ207852
41	46.6	12.6	927	12	AG099877
42	46.4	12.5	465	12	BH254587
43	46.4	12.5	919	12	AG132186
44	46.4	12.5	980	10	BG786274
45	46.2	12.5	812	9	AL537183

ALIGNMENTS

RESULT 1

AG128386
LOCUS Pan troglodytes DNA, clone: PTB-139I13.R, genomic survey sequence.
DEFINITION AG128386
ACCESSION AG128386
VERSION AG128386.1 GI:16657551
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-139I13.R.

ORGANISM

Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-AUG-2001) Asso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chibbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: M13rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. 1090

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-139I13.R"

Fri May 24 10:49:52 2002

BASE COUNT	111 a	34 c	777 g	20 t	148 others
ORIGIN	/sex="male" /cell_type="lymphoblast" /clone_lib="PTB Chimpanzee Male BAC Library" 34 c 777 g 20 t 148 others				
Query Match	13.8%; Score 51.2; DB 12; Length 1090;				
Best Local Similarity	41.7%; Pred. No. 0.31;				
Matches 140; Conservative	0; Mismatches 196; Indels 0; Gaps 0;				
QY	24	ggtgaagggcgggcgccgagccagccagggcgggagggcgggcgggcct	83		
Db	694	GNNNGGGCGCGGGGGGGGGGNGGNGCGCGGGNGGGNGGGGGGGGGGG	753		
QY	84	gtgggaaggtctgggcctggcaggacctgggctgggtctcttggccctgtgtgtgc	143		
Db	754	GGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	813		
QY	144	tttcgcgcaatgctgggtgtctgtacttcgataaacctggagatccctcttttgggcg	203		
Db	814	NGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	873		
QY	204	aatccgggggtagtctctcatcaagactagagtggtgggtgaggaagcttcatacag	263		
Db	874	GGGGGGGNGGGGGGGGGNNNGGGNGAGAATAATGNGCAANAGGNNAGGGGAGGAG	933		
QY	264	gaagcctgtgcgaatgaaggttgccagggaagcatgctgcagaggaactcact	323		
Db	934	AGAGAGCANNAGCANGGAGGAGNGGGGGAAGAAAAAGAGAAGANAGANAG	993		
QY	324	ccgcagaaacacagaaacagagcgagatgagggacg	359		
Db	994	GNAGAAGCAAGAAAGAGAGAGGAGGAGGAGAGG	1029		
RESULT 2					
CNS06LR7/c					
LOCUS	CNS06LR7 941 bp DNA linear GSS 17-JUN-2000				
DEFINITION	T3 end of clone AT00AA016H10 of library ATOAA ATOM from strain CBS 4311 of Saccharomyces servazii genomic survey sequence				

RESULT 2
 CENS06LR7/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 COMMENT

CENS06LR7 941 bp DNA linear GSS 17-JUN-2001
 T3 end of clone AT0AA016H10 of library AT0AA from strain CBS 4311
 of Saccharomyces servazzii, genomic survey sequence.
 AL404537
 AL404537.1 GI:12166153
 GSS.
 Saccharomyces servazzii.
 Saccharomyces servazzii.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 941)
 Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
 Artiguenave,F., Wincker,P. and Gaillardin,C.
 Genomic exploration of the hemiascomycetous yeasts: 7.
 Saccharomyces servazzii
 FEBS Lett. 487 (1); 47-51 (2000)
 20584717
 2 (bases 1 to 941)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrien,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Sautin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 20584711
 3 (bases 1 to 941)
 Genoscope.
 Direct Submission
 Submitted (07-Sep-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Creneau, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen

yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces* *exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces fragilis*, *Kluyveromyces fragilis* var. *marxianus*, *Pichia* *lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia* *lactis* var. *lactis*, *Kluyveromyces marxianus* var. *hansenii*, *Pichia sorbitophila*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See the key words for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES	source	chr	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chr23	chr24	chr25	chr26	chr27	chr28	chr29	chr30	chr31	chr32	chr33	chr34	chr35	chr36	chr37	chr38	chr39	chr40	chr41	chr42	chr43	chr44	chr45	chr46	chr47	chr48	chr49	chr50	chr51	chr52	chr53	chr54	chr55	chr56	chr57	chr58	chr59	chr60	chr61	chr62	chr63	chr64	chr65	chr66	chr67	chr68	chr69	chr70	chr71	chr72	chr73	chr74	chr75	chr76	chr77	chr78	chr79	chr80	chr81	chr82	chr83	chr84	chr85	chr86	chr87	chr88	chr89	chr90	chr91	chr92	chr93	chr94	chr95	chr96	chr97	chr98	chr99	chr100	chr101	chr102	chr103	chr104	chr105	chr106	chr107	chr108	chr109	chr110	chr111	chr112	chr113	chr114	chr115	chr116	chr117	chr118	chr119	chr120	chr121	chr122	chr123	chr124	chr125	chr126	chr127	chr128	chr129	chr130	chr131	chr132	chr133	chr134	chr135	chr136	chr137	chr138	chr139	chr140	chr141	chr142	chr143	chr144	chr145	chr146	chr147	chr148	chr149	chr150	chr151	chr152	chr153	chr154	chr155	chr156	chr157	chr158	chr159	chr160	chr161	chr162	chr163	chr164	chr165	chr166	chr167	chr168	chr169	chr170	chr171	chr172	chr173	chr174	chr175	chr176	chr177	chr178	chr179	chr180	chr181	chr182	chr183	chr184	chr185	chr186	chr187	chr188	chr189	chr190	chr191	chr192	chr193	chr194	chr195	chr196	chr197	chr198	chr199	chr200	chr201	chr202	chr203	chr204	chr205	chr206	chr207	chr208	chr209	chr210	chr211	chr212	chr213	chr214	chr215	chr216	chr217	chr218	chr219	chr220	chr221	chr222	chr223	chr224	chr225	chr226	chr227	chr228	chr229	chr230	chr231	chr232	chr233	chr234	chr235	chr236	chr237	chr238	chr239	chr240	chr241	chr242	chr243	chr244	chr245	chr246	chr247	chr248	chr249	chr250	chr251	chr252	chr253	chr254	chr255	chr256	chr257	chr258	chr259	chr260	chr261	chr262	chr263	chr264	chr265	chr266	chr267	chr268	chr269	chr270	chr271	chr272	chr273	chr274	chr275	chr276	chr277	chr278	chr279	chr280	chr281	chr282	chr283	chr284	chr285	chr286	chr287	chr288	chr289	chr290	chr291	chr292	chr293	chr294	chr295	chr296	chr297	chr298	chr299	chr300	chr301	chr302	chr303	chr304	chr305	chr306	chr307	chr308	chr309	chr310	chr311	chr312	chr313	chr314	chr315	chr316	chr317	chr318	chr319	chr320	chr321	chr322	chr323	chr324	chr325	chr326	chr327	chr328	chr329	chr330	chr331	chr332	chr333	chr334	chr335	chr336	chr337	chr338	chr339	chr340	chr341	chr342	chr343	chr344	chr345	chr346	chr347	chr348	chr349	chr350	chr351	chr352	chr353	chr354	chr355	chr356	chr357	chr358	chr359	chr360	chr361	chr362	chr363	chr364	chr365	chr366	chr367	chr368	chr369	chr370	chr371	chr372	chr373	chr374	chr375	chr376	chr377	chr378	chr379	chr380	chr381	chr382	chr383	chr384	chr385	chr386	chr387	chr388	chr389	chr390	chr391	chr392	chr393	chr394	chr395	chr396	chr397	chr398	chr399	chr400	chr401	chr402	chr403	chr404	chr405	chr406	chr407	chr408	chr409	chr410	chr411	chr412	chr413	chr414	chr415	chr416	chr417	chr418	chr419	chr420	chr421	chr422	chr423	chr424	chr425	chr426	chr427	chr428	chr429	chr430	chr431	chr432	chr433	chr434	chr435	chr436	chr437	chr438	chr439	chr440	chr441	chr442	chr443	chr444	chr445	chr446	chr447	chr448	chr449	chr450	chr451	chr452	chr453	chr454	chr455	chr456	chr457	chr458	chr459	chr460	chr461	chr462	chr463	chr464	chr4
----------	--------	-----	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	------

RESULT	3	
CNS016MU/c		
LOCUS		
DEFINITION		
ACCSSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		

<http://www.eugp.edu.au> ; and Professor


```

/cultivar="Morex"
/db_xref="taxon:4513"
/clone_lib="HVSMEF0010017i"
/hvcDNA0007 (Etiolated and unstressed)
/tissue_type="Seedling root"
/lab_host="TJ121"
/note=Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling roots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give pBluescript
SK(-) cDNA phagemids. These steps were performed in the TJ
Close laboratory at the University of California,
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at The Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To orders
this clone see http://www.genome.clemson.edu/orders Also
see close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html")

```

75 a 35 c 948 g 105 t 33 others

	BASE COUNT	ORIGIN
Query Match	13.4%;	Score 49.4; DB 10; Length 1196;
Best Local Similarity	47.8%;	Pred. No. 0.77;
Matches 143; Conservative	0;	Mismatches 156; Indels 0; Gaps 0;
QY	17	atcgaaagtgaaggcggccgcqggcgtagccaagcccagagggcgagggggac 76
Db	650	AGGGGCGGTGTGCTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 709
QY	77	ggggccttgggaagggtcttgggccttgcaagacctgggctgggtctcttggccctgc 136
Db	710	GGGGGGGGGTGG 769
QY	137	tgtgtctttgcgcaatcgtgggtctgtgactctcgataacctggagatcccttgctt 196
Db	770	TGGGTGGGTGGGGGGGTGTGGTGGGGGGGGTGGGGGGGGGGGGGGGGGGGGTTG 829
QY	197	ttgggcgaatccgggtagtgttcataagactagagtgggggttgagggaagcctt 256
Db	830	GTGTGGGGGTGGGA 889
QY	257	catacagaagcctctgtcgaaaatgaagattgccaggaaaaagcatggcgttcagagg 315
Db	890	GGGAACGGGGGGGGGGGGGTGTGGGGGGGTGGGGAGCTGTGAGGGGGGGGGGGGG 948

RESULT 6

```

AG126129      AG126129      882 bp      DNA      linear      GSS 04-NOV-2001
LOCUS        Pan troglodytes DNA, clone: PTB-136J13.F; genomic sequence.
DEFINITION   AG126129
ACCESSION    AG126129
VERSION      AG126129.1 GI:16655294
KEYWORDS     GSS; GSS (genome survey sequence).
SOURCE       Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            PAC Library clone:PTB-136J13.F.
ORGANISM     Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

```

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and

and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

FEATURES
SOURCE

```
1. .295
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RP11-98"
/clone="BACR39P06"
/notes="end : T7"
```

BASE COUNT	41 a	200 c	5 g	26 t	23 others
ORIGIN					

Query Match	13.1%; Score 48.4; DB 12;	Length 295;
Best Local Similarity	45.6%; Pred. No. 0.88;	
		Models 0.
		Caps 0:

Qy	19	gcgaaggttaaggggcggggccggggcgtagccaagcccgaggggcggagaggggcgg	78
Db	257	GGGGGGGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	198
Qy	79	ggcctgtggaaaggtctgggcctggcaggaacctgggctgggtgtcctctggccctgctg	138
Db	197	GGGGTGTGKGGGGGGGATGAGGGGGGGKGGGGGGKGGKGGKGTCTTAAGKGGGGK	138
Qy	139	tgtctctgcgcgaatgctgggtgtgtgactctcggataacctggagatccctctgtttt	198
Db	137	TGTAGATGGGAGAGTGTGGGGGGGGTGTGGGGGRTKKGSGTKGGTGTGGTGGCG	78
Qy	199	ggcgcaatcccggggttagttgctcatcaagactagagtgaggggtgaggggaagcttc	257
Db	77	CGGACTGCGGGCATGGGTGGAAATGATGANAACCTGATGTGGGTGGGAGGAGTGCC	19

RESULT	10
BE348494/c	
LOCUS	BE348494
DEFINITION	htc9g12.x1 NC1_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3152038 3,
	similar to contains element MSRL repetitive element ; mRNA sequence.
	458 bp linear EST 18-JUL-2000

ACCESSION BE348494
VERSION BE348494.1 GI:9260347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 458)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
CONTACT: Robert Strausberg, Ph.D.
Email: cgapps-r@mail.nih.gov
Tissue Procurement: Christopher Mokaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
CDNA Sequencing by: Washington University Genome Sequencing Center
DNA distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 457.

```

FEATURES
source
Location/Qualifiers
1. .458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3152038"
/clone_lib="NCI CGAP Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"

```

/note="Organ: lung; Vector: pT7730-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-GALP-Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonesIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bentson, Ronald, and M. Fatima Bonaldi."

BASE COUNT	0 a	313 c	22 g	0 t	123 others
ORIGIN					

Query Match	13.0%;	Score 48.2;	DB 10;	Length 458;
Best Local Similarity	34.0%;	Pred. NO. 1.1;		

[illegible]

RESULT 11
A0738512

LOCUS	DEFINITION
LOCUS	DEFINITION

DEFINITION
ACCESSION
VERSION
KEYWORDS

SOURCE ORGANISM

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

REFERENCE

AUTHORS

TITLE

JOURNAL

COCHRANE	MEDLINE	COMMENT
1997	1997	
1998	1998	
1999	1999	
2000	2000	
2001	2001	
2002	2002	
2003	2003	
2004	2004	
2005	2005	
2006	2006	
2007	2007	
2008	2008	
2009	2009	
2010	2010	
2011	2011	
2012	2012	
2013	2013	
2014	2014	
2015	2015	
2016	2016	
2017	2017	
2018	2018	
2019	2019	
2020	2020	
2021	2021	
2022	2022	
2023	2023	
2024	2024	
2025	2025	
2026	2026	
2027	2027	
2028	2028	
2029	2029	
2030	2030	
2031	2031	
2032	2032	
2033	2033	
2034	2034	
2035	2035	
2036	2036	
2037	2037	
2038	2038	
2039	2039	
2040	2040	
2041	2041	
2042	2042	
2043	2043	
2044	2044	
2045	2045	
2046	2046	
2047	2047	
2048	2048	
2049	2049	
2050	2050	
2051	2051	
2052	2052	
2053	2053	
2054	2054	
2055	2055	
2056	2056	
2057	2057	
2058	2058	
2059	2059	
2060	2060	
2061	2061	
2062	2062	
2063	2063	
2064	2064	
2065	2065	
2066	2066	
2067	2067	
2068	2068	
2069	2069	
2070	2070	
2071	2071	
2072	2072	
2073	2073	
2074	2074	
2075	2075	
2076	2076	
2077	2077	
2078	2078	
2079	2079	
2080	2080	
2081	2081	
2082	2082	
2083	2083	
2084	2084	
2085	2085	
2086	2086	
2087	2087	
2088	2088	
2089	2089	
2090	2090	
2091	2091	
2092	2092	
2093	2093	
2094	2094	
2095	2095	
2096	2096	
2097	2097	
2098	2098	
2099	2099	
2100	2100	
2101	2101	
2102	2102	
2103	2103	
2104	2104	
2105	2105	
2106	2106	
2107	2107	
2108	2108	
2109	2109	
2110	2110	
2111	2111	
2112	2112	
2113	2113	
2114	2114	
2115	2115	
2116	2116	
2117	2117	
2118	2118	
2119	2119	
2120	2120	
2121	2121	
2122	2122	
2123	21	

COMMENT

High quality sequence stop: 453.

Seq primer: Sp6

```

    1. 466
    /organism="Trichinella spiralis"
    /db_xref="taxon:6334"
    /clone_lib="Trichinella spiralis"
    /dev_stage="muscle stage larvae"
    /lab_host="puma"

```

/notes=Vector: pCMVSPORT-7.neo; Site_1: NotI; Site_2: SalI
 The library was constructed using mRNA isolated from
 total RNA with oligo-dT cellulose. Total RNA was generated
 from muscle larvae that were isolated from infected rats.
 Larvae were liberated by pepsin/HCl digestion, incubated
 with 1% SDS, treated with RNase and DNase to eliminate
 host nucleic acid contamination, and purified on a Percol
 gradient. The T. spiralis isolate was obtained from
 Dickson Despotnik (Columbia University). The library was
 provided by Dr. Doug Jasmer (djasmer@vetmed.wsu.edu) and
 colleagues at Washington State University. DNA Sequencing
 by: Washington University Genome Sequencing Center St.
 Louis."

28 a	290 c	16 g	22 t	110 others
------	-------	------	------	------------

Query Match 12.9%; Score 47.8; DB 10; Length 466;
Best Local Similarity 40.3%; Pred. No. 1.3;
Matches 100: Conservative 0.4; Moderate 0.4; Radical 0.4;

Query Match	12.9%	Score 47.8;	DB 10;	Length 466;
Best Local Similarity	40.3%	Pred. No. 1.3;		
Matches 100:	Conservative			

Best Local Similarity 40.3%; Pred. No. 1.3;
Matches 100: Conservative 0.1; Microhetero. 1.0;
Score 47.0; DB 10; Length 466;

Consensus	100; Conservative	0; Mismatches	148; Indels	0; Gaps
Qw	10			

[illegible][illegible]

Qy 79 99cctgtt99gaagggtctt99ggcctg9gcaggacctggggtggggtctccttggccctgctg 138

[illegible]

QY 139 tggtctttgcggcaatgctgggtgctgtgactctgggataaacctggagatccctgctctt 198

Qy 259 tacaggaa 266
|||

Db 212 GGNNGGNA 205

RESULT 13
BI415482

LOCUS	BI416482	937 bp	linear	EST 15-AUG-
DEFINITION	hasn001va23f		mRNA	

subtraction cDNA library (hsp) *Pinus sylvestris*/Heterobasidium

amiosum cDNA clone hasp001xa23f, mRNA sequence.
BI416482

VERSION BI416482.1 GI:15187505
KEYWORDS EST.

SOURCE

ORGANISM

pinus sylvestris/Heterobasidium annosum.

pinus sylvestris/Heterobasidium annosum.

REFERENCE

1. Hasegawa, I. and Ohtsuka, K. 1977. Eukaryota; mixed EST libraries. *Journal of Biochemistry* 81: 327-331.

1 (bases 1 to 937)
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (*Pterobasidium annosum*) with

JOURNAL
seedling roots of Scots pine (*Pinus sylvestris*)
Unpublished (2001)

COMMENT

Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,

Sweden

Fri May 24 10:49:52 2002

Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

FEATURES

1. .937
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="has001xa23f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="vector: pT-Adv; Site:1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."
45_c 767 g 21 t 2 others

BASE COUNT
ORIGIN

Query Match	12.9%	Score 47.8	DB 10	Length 937	
Best Local Similarity	47.8%	Pred. No. 1.6			
Matches 139	Conservative	0	Mismatches 152	Indels 0	Gaps 0
QY	25	gtgaggggcccgggcccggggcgtagccaagcccgagggcgagggcgagggcgagggccctg	84		
Db	373	GGGTGGGGGGGGGGGGAGGTGGGTGGGAGAGCGGGGAGGGGGGGGGGGCGGG	432		
QY	85	tgggaagagtcctggcctgcaggaacctggcctgggtctccttggccctgcctgtgtgct	144		
Db	433	CGCGAGGGCGGTCGTGGGAGGGTGGAGGGGGGGGGGGGTAGAGGGCGGGGGAGGAG	492		
QY	145	ttgcggcaatgctgggtgtgtgactctcgtataacctggagatccctgtcttttgggcga	204		
Db	493	GGGGAGGGAGGGCTGGCAGGGGGGTAGGGGGGGGGGAGGGCGGGAGGGGGCGGG	552		
QY	205	atccgggggtagtgtctcatcaagactagagttgggtgtgaggaagagcttcatacagg	264		
Db	553	AGGGGGGGAGGTGGGGAGGGAGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGG	612		
QY	265	aagcctgtcgcgaatgaagagtgtggccaggaaagcatgctgcacagg	315		
Db	613	GAGGGGGGGGGCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	663		

RESULT 14
CNS00G7F/C

CNS00037F	1101 bp	DNA	linear	GSS 03-JUN-1999
CNS00G7F/c				
LOCUS				
DEFINITION				

ACCESSION	VERSION	GT:4951571
AL071731	1	

VERSION.
KEYWORDS

KEYWORDS: *fruit fly*
SOURCE: *fruit fly*

ORGANISM *Drosophila*

Eukaryota;

Pterygota;

Muscomorphology

REFERENCE 1 (bases 1-100)

AUTHORS Genoscope.

TITLE	Direct Subr
--------------	--------------------

JOURNAL
Submitted
25 101 010

BP 161 910

comment
- web : www.
Determinat

COMMENT
determining
collaborat

The BDGP is:

the DDG1 melanogaster

please see

melanogast.

Aaron Mamm

1

JOURNAL

COMMENT

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic Inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES

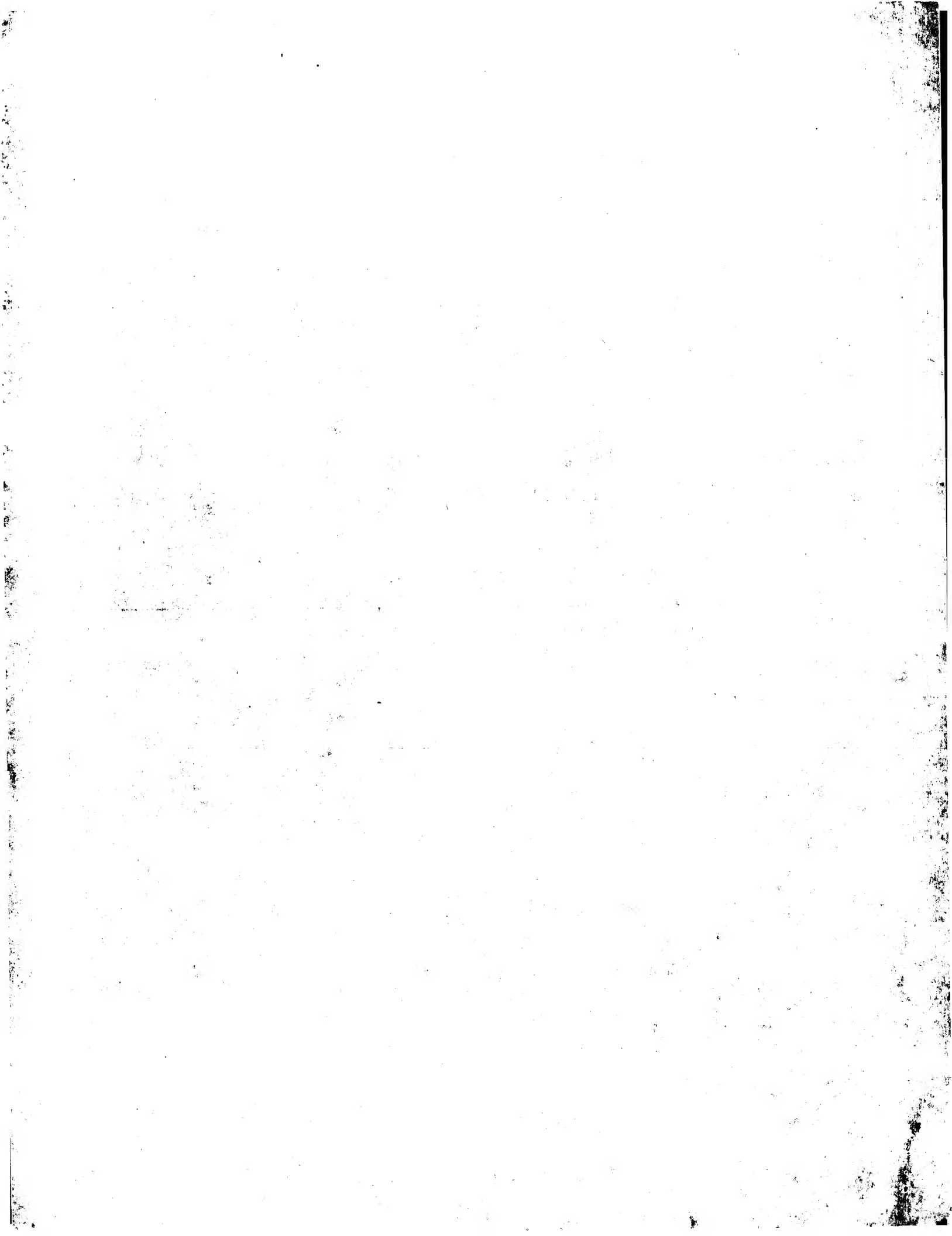
source

Location/Qualifiers
1..922
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0AA030F01"
/clone_lib="AX0AA"
/note="end : 17"

BASE COUNT 216 a 98 C 376 g 138 t 94 others
ORIGIN

Query Match 12.9%; Score 47.6; DB 12; Length 922;
Best Local Similarity 42.4%; Pred. No. 1.7;
Matches 104; Conservative 14; Mismatches 127; Indels 0; Gaps 0;
QY 19 gcaaggtgagggcgccgcccgcgtagcccaagcggcgaggggggcggg 78
Db 327 GCKAGGGGSKSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 78
QY 79 ggcctgtggaagggtctggccctgcagaccctggctgggtctctcttggccctgtg 138
Db 387 GGG 446
QY 139 tgtgttttcgcgcaatgctgggtgctgtgactctcgataacctggagatccctgtttt 198
Db 447 GGG 506
QY 199 gggcaatccgggtagttactatcaagactagagtgagggtgaggaaggcttca 258
Db 507 GCGKSGGGGGGGRGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGG 566
QY 259 tacag 263
Db 567 TATWG 571

Search completed: May 23, 2002, 23:06:13
Job time: 7529 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:20:45 ; Search time 121.39 seconds
(without alignments)
748.698 Million cell updates/sec

Title: US-09-696-686-51

Perfect score: 370

Sequence: 1 tgcgattcccgcaaatgc.....atgaggagccctgcggcc 370

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	50.4	13.6	7218	1	US-08-232-463-14
C 2	39.2	10.6	319	4	US-09-165-264-8
C 3	39	10.5	152331	3	US-09-128-155-16
C 4	37.6	10.2	318	4	US-09-165-264-12
C 5	37.6	10.2	320	4	US-09-165-264-13
C 6	37.2	10.1	320	4	US-09-165-264-14
C 7	37	10.0	320	4	US-09-165-264-7
C 8	36	9.7	4405	1	US-07-885-972A-3
C 9	36	9.7	4405	2	US-08-745-880-3
C 10	36	9.7	4405	2	US-08-480-382-3
C 11	35.8	9.7	4488	4	US-08-406-030A-3
C 12	35.6	9.6	320	4	US-09-165-264-11
C 13	35.4	9.6	4046	1	US-07-793-961A-1
C 14	35.4	9.6	4046	1	US-08-240-357-1
C 15	34.8	9.4	374	2	US-08-370-156-24
C 16	34.8	9.4	1215	2	US-08-370-156-26
C 17	34.8	9.4	2256	2	US-08-318-826A-5
C 18	34.8	9.4	2256	2	US-08-370-156-1
C 19	34.8	9.4	2256	3	US-08-814-095-1
C 20	34.8	9.4	3016	2	US-08-318-826A-7
C 21	34.8	9.4	3016	2	US-08-370-156-5
C 22	34.8	9.4	3016	3	US-08-814-095-5
C 23	34.8	9.4	3096	2	US-08-318-826A-6
C 24	34.8	9.4	3096	2	US-08-370-156-3
C 25	34.8	9.4	3096	3	US-08-814-095-3
C 26	34.2	9.2	53525	3	US-08-658-136-2
C 27	34.2	9.2	53577	3	US-08-658-136-1

28	34	9.2	1737	2	US-08-750-703-2	Sequence 2, Appli
29	34	9.2	8779	2	US-08-750-703-4	Sequence 4, Appli
C 30	33.8	9.1	2689	2	US-08-985-090-1	Sequence 1, Appli
C 31	33.8	9.1	2689	3	US-09-165-543-1	Sequence 11, Appli
32	33.6	9.1	5663	4	US-09-499-884-11	Patent No. 5215881
33	33.4	9.0	2721	6	5215881-2	Sequence 1, Appli
C 34	33.4	9.0	4257	2	US-08-690-473-1	Sequence 1, Appli
C 35	33.4	9.0	4257	4	US-09-259-821A-1	Sequence 1, Appli
C 36	33.4	9.0	4257	4	US-08-843-659-1	Sequence 1, Appli
C 37	33.4	9.0	4425	2	US-08-749-169A-1	Sequence 1, Appli
C 38	33.4	9.0	4425	2	US-09-130-032A-1	Sequence 1, Appli
39	33.4	8.9	12001	1	US-08-458-588A-11	Sequence 11, Appli
40	33	8.9	1702	1	US-08-459-586-9	Sequence 9, Appli
41	33	8.9	1702	2	US-08-282-696-9	Sequence 1, Appli
42	33	8.9	8438	1	US-07-945-283-1	Sequence 1, Appli
43	33	8.9	18994	1	US-08-459-586-4	Sequence 4, Appli
44	33	8.9	18994	2	US-08-282-696-4	Sequence 4, Appli
45	32.8	8.9	640	2	US-08-835-099A-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Fri May 24 10:49:52 2002

```

Query Match          13.6%; Score 50.4; DB 1; Length 7218;
Best Local Similarity 0.9%; Pred. No. 0.00045;
Matches 3; Conservative 214; Mismatches 135; Indels 0; Gaps 0;

QY      6   ttgccagcaaatcgaaagtggagggggcgcccggggcgtagccaaagccccagggg 65
       | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1440 TGTGACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1381

QY     66   cgggaaggggcgggcctgtggaaggcttgccctgcaggacctggcgtgggtctc 125
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321

QY    126   ctTggccctgtgtgtcttgcggcaatcgtggctgctgactctcgataaacctgga 185
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1361

QY    186   gatccctgttttggcgaaatccggggggttagttgctcatcaaagactagagggtgg 245
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1360 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201

QY    246   aggaagccttcacagggaagcctctgcgaaatgaagagttTggccagggaagcatgg 305
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141

QY    306   cgtgcagagaactcattccgcagaaaaccacagaaacagaggcagatgagga 357
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1089


RESULT 2
US-09-165-264-8
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinavagamoorthy, Thuraiyarah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8
```

```

Query Match          10.6%;   Score 39.2;   DB 4;   Length 319;
Best Local Similarity 58.6%;   Pred. No. 0.15;
Matches 68;   Conservative 0;   Mismatches 48;   Indels 0;   Gaps 0;

Qy      19  qcqaagtgagggcgcgggcgcgccgtagccaagcccgagggcgqggagggcgggcg 78
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      198 ggggggggggggggggggggggggggggggggggggggggggggggggg 257

Qy      79  ggcctgtgtgggaaggtctgtgggcctggcaggaacctgggctgggtctcttgccct 134
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      258 ggggggggggggggggggggggggggggggggggggggggggggggggg 313

RESULT      3
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155

```

```

; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match          10.5%;   Score 39;    DB 3;   Length 152331;
Best Local Similarity 63.2%; Pred.No. 0.62;
Matches 60; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY      27 gaggggccccgcgcccggttagccaagcccgaggcgggagggcgggccctgtg 86
        ||||| |||||| | | | | | | | | | | | | | | | | | | | | |
DB      22204 GCGGCGGCGCGCGCGGGGGGGGGCGGGCGGGCGGGGGGTCTCGGGGTCCG 22145

QY      87 ggaagggtctggcctggcaggacctgggctgggg 121
        || ||| | | | | | | | | | | | | | | | | | | | | | | | |

DB      22144 GGCGCGGCGCGCGCGGGGGGGGGGGGGTCTCGCTGCCG 22110


RESULT      4
US-09-165-264-12
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12
```

```

Query Match      10.2%; Score 37.6; DB 4; Length 318;
Best Local Similarity 56.5%; Pred. No. 0.38;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps
0

QY 19 gcgaaaggttagggggccgcgcgcgttagccaagcccgagggcgggagggggcg 78
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 ggggggggggggggggggggggggggggggggggggggggggggggggg 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 79 ggcctgtgggaagggtctgtggcctggcagacactgagctggggtctccttggccctgctg 138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 ggggggggggggggggggggggggggggggggggggggggggggggggg 306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 139 ttgt 142
      ||| |||
Db. 307 ttgtg 310
      ||| |||

RESULT 5
US-09-165-264-13
; Sequence 13, Application US/09165264
; Patent No. 6197510

```

```
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13
```

```
Query Match
Best Local Similarity 10.2%; Score 37.6; DB 4; Length 320;
Matches 64; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 19 gcgaaggtgaggggcccgcgcgtagcgaagcccgagggcgagggcgcg 78
Db 211 ggggggggggggggggggggggggggggggggggggggggggggggg
QY 79 ggcctgtggaagggctctggccctggcagacctggggtctctcc 126
Db 271 ggggggggggggggggggggggggggggggggggggggggggggggg
RESULT 6
US-09-165-264-14
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14
```

```
Query Match
Best Local Similarity 10.1%; Score 37.2; DB 4; Length 320;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 19 gcgaaggtgaggggcccgcgcgtagcgaagcccgagggcgagggcgcg 78
Db 198 ggggggggggggggggggggggggggggggggggggggggggggggg
QY 79 ggcctgtggaagggctctggccctggcagacctggggtctctctcc 138
Db 258 ggggggggggggggggggggggggggggggggggggggggggggggg
QY 139 tg 140
Db 318 ag 319

RESULT 7
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
```

```
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7
```

```
Query Match
Best Local Similarity 10.0%; Score 37; DB 4; Length 320;
Matches 112; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 19 gcgaaggtgaggggcccgcgcgtagcgaagcccgagggcgagggcgcg 78
Db 69 ggggggggggggggggggggggggggggggggggggggggggggggg
QY 79 ggcctgtggaagggctctggccctggcagacctggggtctctctcc 138
Db 129 ggggggggggggggggggggggggggggggggggggggggggggggg
QY 139 tgccttttcgcgaatgctgggtgctgacctctcgataacctggagatccctgctttt 198
Db 189 ggggggggggggggggggggggggggggggggggggggggggggggg
QY 199 gggcgaatccggggtagttgctcatcaagactagaggtgagggcgagggcgct 255
Db 249 ggggggggggggggggggggggggggggggggggggggggggggggg
RESULT 8
US-07-885-972A-3/c
; Sequence 3, Application US/07885972A
; Patent No. 5460950
; GENERAL INFORMATION:
; APPLICANT: Barr, Phillip J.
; APPLICANT: Brake, Anthony J.
; APPLICANT: Kaufman, Knadal J.
; APPLICANT: Tekamp-Olson, Patricia
; APPLICANT: Wasley, Louise
; APPLICANT: Wong, Polly A.
; TITLE OF INVENTION: Expression of PACE in Host Cells and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson & Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,972A
; FILING DATE: 19920520
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/621,092
; FILING DATE: 26-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,859
; FILING DATE: 29-NOV-1990
```

ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/745,880
FILING DATE: 08-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,382
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 07/885,972

```

RESULT 10
US-08-480-382-3/c
; Sequence 3, Application US/08480382
; Patent No. 5986079
; GENERAL INFORMATION:
; APPLICANT: Barr, Philip J.
; APPLICANT: Brake, Anthony J.
; APPLICANT: Kaufman, Rnadal J.
; APPLICANT: Tekamp-Olson, Patricia
; APPLICANT: Wasley, Louise
; APPLICANT: Wong, Polly A.
; TITLE OF INVENTION: Expression of PACE in Host Cells and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson & Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,382

```

```

RESULT 11
US-08-406-030A-3
; Sequence 3, Application US/08406030A
; Patent No. 6370989
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Hauge, Brian M.
; APPLICANT: Selden, Richard F
; TITLE OF INVENTION: Protein Production and Delivery
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millcia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

RESULT 12
US-09-165-264-11
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analyses
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
US-09-165-264-11

```

```

Query Match      9.6%; Score 35.6; DB 4; Length 320;
Best Local Similarity 44.6%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 140; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 19 gcgaagtgagggcgccgagggcgtagcccaagcccgagggcgagggcgagggcg 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 79 ggcctgtggaaggtctgggcccctggcagacactggctgggctccttgcctctgtg 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 139 tggcttgcgcaactgctgggtgctgactctcgataaactggagatccctgtttt 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 199 gggcgaaatccggggttagttgctcatcaagactagaggtggggtggaagccttca 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 259 tacagaagcctctcgcgaataaagaagtggtggccaggaagacatgcatgcagagaaac 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 ggggggggggggggggggggggggggggggggggggggggggggggggggggg 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 319 tcactccgcagaaa 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 agatttgaagaaa 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-07-793-961A-1
; Sequence 1, Application US/07793961A
; Patent No. 5334521
; GENERAL INFORMATION:
; APPLICANT: Yoshihiro Ishikawa
; TITLE OF INVENTION: Cloning and Character-
; TITLE OF INVENTION: ization of a Cardiac Adenylyl Cyclase
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Alan M. Gordon
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street,
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII from DW4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/793.961A
; FILING DATE: 19911118
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,705
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2719
; TELEFAX: 203 321 2971
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4046 base pairs listed
; TYPE: NUCLEIC ACID

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-793-961A-1

Query Match      9.6%; Score 35.4; DB 1; Length 4046;
Best Local Similarity 61.3%; Pred. No. 2.3; Indels 0; Gaps 0;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 27 gagggggcgggcgccgagggcgtagcccaagcccgagggcgagggcgagggcgctgtg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 ggaagggtctgggcccctggcgagacacctgggctgg 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GGGCGGGGTCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-240-357-1
; Sequence 1, Application US/08240357
; Patent No. 5578481
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Yoshihiro
; TITLE OF INVENTION: Cloning and Characterization of a
; TITLE OF INVENTION: Cardiac Adenylyl Cyclase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,357
; FILING DATE: 10-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,705-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3244
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4046 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 131...3625
US-08-240-357-1

```

```

Query Match      9.6%; Score 35.4; DB 1; Length 4046;
Best Local Similarity 61.3%; Pred. No. 2.3;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 27 gagggggcgggcgccgagggcgtagcccaagcccgagggcgagggcgagggcgctgtg 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 ggaagggtctgggcccctggcgagacacctgggctgg 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 94 GCGGGGTCCGGCGCGCGGAGCGGGCGCG 126
|| ||||| ||| | | ||| ||||| ||

RESULT 15

US-08-370-156-24/c
; Sequence 24, Application US/08370156
; Patent No. 5932780
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTICHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethlington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,156
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-307 (Mulford)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-370-156-24

Query Match 9.4%; Score 34.8; DB 2; Length 374;
Best Local Similarity 57.3%; Pred. No. 2;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
Qy 6 ttgcccagcaaatcgagggtgaggggcgggcgggcggtagccaagcccgagggg 65
||| | ||||| | ||| | | | | | ||| |||||
Db 322 TTGCAGCAGAGGCGGAGGCGGCGGAGGAGCTGAGACATGCCAGAGGAC 263
||| | ||||| | ||| | | | | | ||| |||||
Qy 66 cggagggggcgggcctgtgggaaggtctggccctggcaggaacctggg 115
||| | ||||| | ||| | | | | | ||| |||||
Db 262 CGGAGAGCGCGGAGGAGCTCGGGGTGGGGATGGGCAGAGTCTGGG 213

Search completed: May 24, 2002, 00:21:22
Job time: 11748 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 24, 2002, 00:24:32 ; Search time 525.6 Seconds
(without alignments)
1208.635 Million cell updates/sec

Title: US-09-696-686-51
Perfect score: 370
Sequence: 1 tgcgattgccgcaaatgc.....atgagagacgcctgccgccc 370

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				N_Geneseq_032802.*			
				1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*		
				2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*		
				3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*		
				4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*		
				5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*		
				6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*		
				7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*		
				8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*		
				9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*		
				10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*		
				11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*		
				12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*		
				13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*		
				14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*		
				15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*		
				16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*		
				17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*		
				18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*		
				19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*		
				20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*		
				21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*		
				22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*		
				23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*		
				24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	370	100.0	370	22	AAS05285
2	42.4	11.5	4362	22	AA05795
3	41.8	11.3	1337	20	AAZ17263
4	41.2	11.1	972	22	AAAF81304
5	41.2	11.1	972	23	AAAF81304
6	40	10.8	38186	20	AAZ32028
7	40	10.8	38186	22	AAZ32028
8	39.6	10.7	114953	20	AAZ33491
9	39	10.5	3198	20	AAZ02974

C 10	38.6	10.4	10828	22	AA07171	Human reproductive
C 11	38.4	10.4	357	22	AA081967	Human polynucleoti
C 12	38.2	10.3	1000	21	AAA02484	Human colon cancer
C 13	37.6	10.2	318	21	AAA38184	Primer used in the
C 14	37.6	10.2	320	21	AAA38185	Primer used in the
C 15	37.4	10.1	921	23	AAAF7438	DNA encoding novel
C 16	37.4	10.1	1286	24	AB199656	Mouse ischaemic co
C 17	37.4	10.1	1315	20	AAZ41312	Human normal ovari
C 18	37.4	10.1	21747	22	AAZ41312	Human immune/haema
C 19	37.2	10.1	320	22	AAZ76413	Primer used in the
C 20	37.2	10.1	419	23	AAZ38186	DNA encoding novel
C 21	37.2	10.1	154746	24	AAZ38186	Human herpesvirus
C 22	37.2	10.1	154746	24	AAZ38186	Human colon cancer
C 23	37	10.0	2152	22	AAH15432	Human ovarian tumo
C 24	36.8	9.9	1593	21	AAZ02504	M. tuberculosis an
C 25	36.8	9.9	2188	20	AAZ77506	Corn Wuschel (WUS)
C 26	36.6	9.9	1166	20	AAZ19440	DNA encoding novel
C 27	36.6	9.9	1166	20	AAZ19440	Human immune/haema
C 28	36.4	9.8	513	22	AAZ02223	Human polynucleoti
C 29	36.4	9.8	792	23	AAZ02223	Human gene express
C 30	36.4	9.8	3485	23	AAZ02223	Human polynucleoti
C 31	36.4	9.8	10468	22	AAZ77961	Bovine-derived DNA
C 32	36.2	9.8	434	22	AAZ77961	Mouse ischaemic co
C 33	36.2	9.8	1312	20	AAZ17264	Human secreted pro
C 34	36	9.7	349	22	AAZ17264	PACE composite seq
C 35	36	9.7	1008	22	AAH21004	Human PACE coding
C 36	36	9.7	1045	24	AAH21004	Human PACE-like pr
C 37	36	9.7	1529	21	AAZ199807	Human PACE encodin
C 38	36	9.7	4405	13	AAZ02544	CDNA encoding nove
C 39	36	9.7	4405	16	AAZ02544	Human immune/haema
C 40	36	9.7	4405	20	AAZ02544	Human METH2 relate
C 41	36	9.7	4405	21	AAZ02544	Human METH2 relate
C 42	36	9.7	6288	22	AAZ02544	Human METH2 relate
C 43	36	9.7	7213	20	AAZ02544	Human METH2 relate
C 44	36	9.7	8670	22	AAZ02544	Human METH2 relate
C 45	36	9.7	8670	22	AAZ02544	Human METH2 relate

ALIGNMENTS

RESULT 1	
AAS05285	
ID	AAS05285 standard; DNA; 370 BP.
AC	AAS05285;
DT	07-SEP-2001 (first entry)
DE	T243 gene homologous sequence #2 generated by PCR.
DE	Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW	Transgenic animal; knockout mouse; triplet repeat expansion;
KW	fragile X syndrome; Huntington's disease; mouse; ds.
OS	Mus sp.
PN	WO200130798-A1.
PD	03-MAY-2001.
XX	26-OCT-2000; 2000WO-US29382.
XX	26-OCT-1999; 99US-0161488.
XX	(DELT-) DELTAGEN INC.
PI	Klein R, Matthews W, Moore M, Allen KD;
XX	WPI; 2001-300473/31.
DR	Novel transgenic animals useful as animal model for characterization of
PT	function of a gene encoding trinucleotide repeat proteins (TRPs),

XX	WO200121791-A2.
XX	
NN	29-MAR-2001.
XX	
DD	25-SEP-2000; 20
XX	23-SEP-1999;
XX	(GLAX) GLAXO G
XX	Clare JJ, Plum
PPI	WPI; 2001-26607
XX	P-PSDR: AAR739
DR	

(GLAXO) GLAXO GROUP LTD.
 Clare JJ, Plumpton M, Moss FJ, Sanseau P;
 MPI: 2001-266074/27.
 P-PSDB: AAB73982.
 Novel stargazin-like polypeptides capable of modulating the steady
 inactivation of an alpha pore-forming subunit of a voltage-gated
 calcium channel, for treating central nervous system disorders -
 Claim 7. Page 45-47: 48pp: English.

Sequence 972 BP: 193 A; 291 C; 266 G; 222 T; 0 other:
migraine, headache, bipolar disorder, unipolar depression, anxiety,
Parkinson's disease, cognitive disorders, ophthalmic diseases,
neuromuscular disorders and tinnitus. Nucleic acids encoding these
polypeptides are useful in the production of the polypeptides, and as
primers.

```

Query Match          11.1%   Score 41.2;   DB 22;   Length 972;
Best Local Similarity 49.1%;   Pred. No. 0.22;
Database 109: conservative 0; Mismatches 113; Indels 0; Gaps
primers.
Sequence 972 BP; 193 A; 291 C; 266 G; 222 T; 0 other;

```

	RESULT	5
	AAS80100	
ID	AAS80100 standard; cDNA; 972 BP.	
XX	XX	
XX	AAS80100;	
XX	AC	
XX	XX	
XX	DT	
XX	DT	
XX	(first entry)	
	13-FEB-2002	
	DNA encoding novel human diagnostic protein #15904.	

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG15913.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID No 15904; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 972 BP; 193 A; 291 C; 266 G; 222 T; 0 other;

Query Match 11.1%; Score 41.2; DB 23; Length 972;
Best Local Similarity 49.1%; Pred. No. 0.22;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 3 cgaatgccagcaaatgcaagtgtagggcgggcgccgagcgtgacccaagcccgag 62
Db 45 cgaagccttgaaagggccttgtagtgaggcgccagcccgagcgtgctgctgt 104
QY 63 gggcgggagggcgggcgccctgctggaaggtctggtgagcagccctggcggtg 122
Db 105 gggagcgtggcagctagtcacagcaacggcggcgaagatgagtcgtcgaggaa 164
QY 123 ctccctggccctgctgtgtgttcgggcaatgctggtgctgactctcggaacact 182
Db 165 ggcctgacctgctgagcagctgtctgtgctgctggtggtgctgctggtatcgc 224
QY 183 ggaatccctcttttggcgcaatccgggggttagttctcat 224
Db 225 ggtcagcaccgactactggtgctgtacctggagagaggtgtgat 266

RESULT 6
AAZ32028
ID AAZ32028 standard; DNA; 38186 BP.
XX
AC AAZ32028;
XX
DT 10-JAN-2000 (first entry)
XX
DE Human METH1 related EST AC004449.
XX
KW Human: METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.
XX
OS Homo sapiens.
XX
PN WO9937660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
XX 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PI
XX IrueLa-Arispe L, Hastings GA, Ruben SM;
XX WPI; 1999-590684/50.
XX
XX New isolated metalloprotease thrombospondin polypeptides, useful for
XX treating hyperproliferative disorders, cancers or autoimmune disorders
XX
XX Disclosure: Page 363-387; 457pp; English.
XX
XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
XX metalloprotease thrombospondin (METH) proteins METH1 and METH2
XX respectively. METH1 and METH2 have been found to be potent inhibitors of
XX angiogenesis both in vitro and in vivo. They can be used for treating
XX cancer and other disorders related to angiogenesis including abnormal
XX wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX endometrial bleeding disorders, rheumatoid arthritis, psoriasis,
XX macula degeneration, haemangiomas, and arterial-venous malformations.
XX They may be useful in treating deficiencies or disorders of the immune
XX system, by activating or inhibiting the proliferation, differentiation,
XX or mobilisation (chemotaxis) of immune cells. The etiology of these
XX immune deficiencies or disorders may be genetic, somatic, such as
XX cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
XX toxins), or infectious. They can also be used to treat inflammatory
XX conditions, both chronic and acute conditions. The products can also be
XX used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
XX AAY49511 represent sequences given in the exemplification of the present
XX invention.
XX
XX Sequence 38186 BP; 7571 A; 11503 C; 12193 G; 6919 T; 0 other;

Query Match 10.8%; Score 40; DB 20; Length 38186;
Best Local Similarity 55.9%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 8 gccacgcaaatcgaaagtgagggcgccgagcgcggcgtagccaagcccgagggcg 67
Db 2699 ggccggggatggggatgggagtgaggcgccgggagtgaggcgggcggtggcgccg 2758

RESULT	9	
AAAX02974/c		
ID	AAAX02974	standard; DNA; 3198 BP.
XX		
XX	AAAX02974;	
XX		
XX	22-JUN-1999	(first entry)
DT		
XX		
XX	Human IL-1ra BAC	contiguous DNA sequence 19.
DE		
XX		
XX	Tango-77; human; IL-1ra;	cytokine superfamily; inflammation; inhibition;
KW	Interleukin-1 receptor; IL-1R;	regulation; asthma; rheumatoid arthritis;
KW	chronic myelogenous leukaemia;	psoriasis; inflammatory bowel disease;
KW	growth factors; treatment;	IL-1 receptor complex; BAC; ss.
XX		
XX		
OS	Homo sapiens.	
XX		
PN	WO9906426-A1.	
XX		
PD	11-FEB-1999.	
XX		
PF	03-AUG-1998;	98WO-US16102.
XX		
PR	02-JUL-1998;	98US-0091650.

RESULT	10	
AAL07171/c		
ID	AAL07171 standard; DNA; 10828 BP.	
XX		
XX		
AC	AAL07171;	
XX		
DT	21-NOV-2001 (first entry)	
XX		
DE	Human reproductive system related antigen DNA SEQ ID NO: 9859.	
XX		
KW	Human; reproductive system related antigen; reproductive system disorder;	
KW	cancer; gene therapy; ds.	
XX		
XX		
OS	Homo sapiens.	
XX		
PN	WO200155320-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01339.	
XX		
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	

PR	28-JUN-2000;	2000US-0214886.	PR	20-OCT-2000;	2000US-0241786.
PR	30-JUN-2000;	2000US-0215135.	PR	20-OCT-2000;	2000US-0241787.
PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000;	2000US-0216880.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	11-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246610.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246611.
PR	22-AUG-2000;	2000US-0226681.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226868.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249208.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249209.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249210.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249214.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249216.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249218.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249266.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0232081.	PR	17-NOV-2000;	2000US-0249299.
PR	12-SEP-2000;	2000US-0231968.	PR	17-NOV-2000;	2000US-0249300.
PR	14-SEP-2000;	2000US-0232397.	PR	01-DEC-2000;	2000US-0250160.
PR	14-SEP-2000;	2000US-0232398.	PR	01-DEC-2000;	2000US-0250160.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232400.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232401.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0233063.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251856.
PR	21-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251868.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251869.
PR	21-SEP-2000;	2000US-0234274.	PR	08-DEC-2000;	2000US-0251989.
PR	25-SEP-2000;	2000US-0234397.	PR	08-DEC-2000;	2000US-0251990.
PR	25-SEP-2000;	2000US-0234998.	PR	11-DEC-2000;	2000US-0254097.
PR	26-SEP-2000;	2000US-0234998.	PR	05-JAN-2001;	2001US-0259678.
PR	26-SEP-2000;	2000US-0235484.	XX	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	27-SEP-2000;	2000US-0235836.	PI	WPI; 2001-465570/50.	
PR	29-SEP-2000;	2000US-0236327.	XX	Isolated nucleic acid molecule encoding a reproductive system antigen -	
PR	29-SEP-2000;	2000US-0236367.	XX	is used in preventing, treating or ameliorating a medical condition -	
PR	29-SEP-2000;	2000US-0236368.	XX	Disclosure; SEQ ID NO 9859; 1297pp + Sequence Listing; English.	
PR	29-SEP-2000;	2000US-0236369.	XX	The present invention provides the protein and coding sequences of a	
PR	29-SEP-2000;	2000US-0236370.	XX	number of human reproductive system related antigens. These can be used	
PR	02-OCT-2000;	2000US-0236802.	XX	in the prevention and treatment of reproductive system disorders,	
PR	02-OCT-2000;	2000US-0237037.	XX	including cancer. The present sequence is a genomic sequence encoding a	
PR	02-OCT-2000;	2000US-0237038.	XX	protein of the invention.	
PR	02-OCT-2000;	2000US-0237039.	XX	Sequence 10828 BP; 2518 A; 3387 C; 2921 G; 2002 T; 0 other;	
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	XX		
PR	13-OCT-2000;	2000US-0239937.	XX		
PR	20-OCT-2000;	2000US-0240960.	XX		
PR	20-OCT-2000;	2000US-0241221.	XX		
PR	20-OCT-2000;	2000US-0241785.	SQ		

CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 1000 BP; 21 A; 53 C; 574 G; 22 T; 330 other;

Query Match 10.3%; Score 38.2; DB 21; Length 1000;
Best Local Similarity 36.3%; Pred. No. 1.3;
Matches 91; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 24 ggtgagggggcgccgcccggcgtagccaagcccgagggcgagggcgccgct 83
DB 361 ggggngngggngnggggngnnnggggggggagngnnngngnggggng 420
QY 84 gtgggaagggtctggccctggcagaccctggctgggtgtctctctgcccctgtgtgc 143
DB 421 gggggngngcngngngggggnggggngngngnggagcgngggngcngngngn 480
QY 144 ttgcggcaatgctgggtgctgtgactctcggaataacctggagatccctgtttggcg 203
DB 481 naggncnggggngngngngnggnggnggngngngngngngngngngngng 540
QY 204 aatccggggtagtgcctcatcaagactagagtggtgggtggaggaagcttcatacag 263
DB 541 ngggnggggng 600
QY 264 gaagcctctgt 274
DB 601 gangggngng 611

RESULT 13
AAA38184
ID AAA38184 standard; DNA; 318 BP.
XX
AC AAA38184;
XX
DT 01-SEP-2000 (first entry)
XX
DE Primer used in the analysis of a BVDV genome fragment.
XX
KW Primer; bovine viral diarrhea virus; BVDV; nucleic acid analysis;
KW diagnosis; pathological organism; detect; ss.
XX
OS Bovine viral diarrhea virus.
XX
PN WO200020628-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00915.
XX
PR 01-OCT-1998; 98US-0165264.
XX
PA (BIOI-) BIO-ID DIAGNOSTIC INC.
XX
PI Vinayagamoorthy T;
XX
DR WPI; 2000-303800/26.
XX
PT Nucleic acid analysis methods for simultaneously analysing multiple
PT nucleic acid regions for diagnosis and differentiation of pathological
PT organisms comprises sequencing the nucleic acids in the reaction
PT mixture -
XX
PS Example 2; Page 23; 36pp; English.
XX
CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhea virus (BVDV) genome. The primer is used to
CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for

CC two nucleic acid target sequences. The methods are used for
CC simultaneously analysing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the size of the amplified
CC region to be increased. The technique is especially useful when the usual
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subsets of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC predisposition to a disease.
XX
SQ Sequence 318 BP; 3 A; 3 C; 306 G; 6 T; 0 other;

Query Match 10.2%; Score 37.6; DB 21; Length 318;
Best Local Similarity 56.5%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 19 gcgaagggtgagggggcgccgcccggcgtagccaagcccgagggcgagggcgcg 78
DB 187 ggg 246
QY 79 ggcctgtgggaagggtctggccctggcagacctggtggtctctcttggccctgctg 138
DB 247 ggg 306
QY 139 tgtg 142
DB 307 tgtg 310

RESULT 14
AAA38185
ID AAA38185 standard; DNA; 320 BP.
XX
AC AAA38185;
XX
DT 01-SEP-2000 (first entry)
XX
DE Primer used in the analysis of a BVDV genome fragment.
XX
KW Primer; bovine viral diarrhea virus; BVDV; nucleic acid analysis;
KW diagnosis; pathological organism; detect; ss.
XX
OS Bovine viral diarrhea virus.
XX
PN WO200020628-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00915.
XX
PR 01-OCT-1998; 98US-0165264.
XX
PA (BIOI-) BIO-ID DIAGNOSTIC INC.
XX
PI Vinayagamoorthy T;
XX
DR WPI; 2000-303800/26.
XX
PT Nucleic acid analysis methods for simultaneously analysing multiple
PT nucleic acid regions for diagnosis and differentiation of pathological
PT organisms comprises sequencing the nucleic acids in the reaction
PT mixture -
XX
PS Example 2; Page 23; 36pp; English.
XX
CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhea virus (BVDV) genome. The primer is used to
CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:18:14 ; Search time 3953.4 Seconds
(without alignments)
1958.521 Million cell updates/sec

Title: US-09-696-686-51
Perfect score: 370
Sequence: 1 tgcgattgccagcaaatgc.....atgaggagccctgccgccc 370

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
1797656 seqs, 10463268293 residues

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

GenEmbl: *
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_Other.*
33: em_htgo_inv.*

```

ALIGNMENTS

RESULT 1

AC074220/c

LOCUS

DEFINITION

AC074220

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match	Length	DB	ID	Description
1	AC074220	99.6	368.4	1	1	AC074220 Mus muscu
2	AC102804	50.8	50.8	2	2	AC102804 Homo sapi
3	AC106591	13.7	53178	2	3	AC106591 Rattus no
4	I66494	50.6	13.7	2	4	I66494 Sequence 14
5	OCU05204	49.2	13.6	6	5	OCU05204 Oryctolagus
6	OCU05204	49.2	13.3	1396	6	OCU05204 Oryctolagus
7	AC023197	13.2	141892	2	7	AC023197 Mus muscu
8	AC094865	48.8	13.2	236301	2	AC094865 Rattus no
9	AC095701	48.2	13.0	163525	2	AC095701 Rattus no
10	AC095225	47.8	12.9	148921	2	AC095225 Rattus no
11	AC100281	47.6	12.9	46742	2	AC100281 Mus muscu
12	AC093968	47.2	12.9	199536	2	AC093968 Rattus no
13	AC095175	47.2	12.8	107781	2	AC095175 Rattus no
14	AC079431	47.2	12.8	306695	2	AC079431 Mus muscu
15	AC100680	47.2	12.7	57787	2	AC100680 Mus muscu
16	AC094597	47.2	12.7	192182	2	AC094597 Rattus no
17	AC084799	46.8	12.7	303091	2	AC084799 Mus muscu
18	AC100569	46.8	12.6	80692	2	AC100569 Mus muscu
19	AC022588	46.6	12.6	83296	2	AC022588 Homo sapi
20	AC101154	46.6	12.6	80291	2	AC101154 Mus muscu
21	AC021929	46.6	12.6	81767	2	AC021929 Homo sapi
22	AC016280	46.4	12.6	144979	2	AC016280 Homo sapi
23	AC090865	46.4	12.5	51119	2	AC090865 Homo sapi
24	AC097363	46.4	12.5	68226	2	AC097363 Rattus no
25	HS117M_1	46.4	12.5	110000	2	HS117M_1 Continuation (2 of
26	HS21C101	46.2	12.5	340000	9	HS21C101 Homo sapi
27	AC022595	46.2	12.5	82897	2	AC022595 Homo sapi
28	AC095300	46.2	12.5	145070	2	AC095300 Rattus no
29	AC027682	46.2	12.5	205914	2	AC027682 Homo sapi
30	AC087563	46.2	12.5	298166	2	AC087563 Homo sapi
31	CNS06EET	46.2	12.4	756	11	CNS06EET AL395019 T3 end of
32	AC027483	46.2	12.4	75144	2	AC027483 Homo sapi
33	AC107595	46.2	12.4	76101	2	AC107595 Rattus no
34	AC106169	45.8	12.4	159760	2	AC106169 Rattus no
35	AC087337	45.8	12.4	23191	2	AC087337 Homo sapi
36	AC101222	45.8	12.4	57971	2	AC101222 Mus muscu
37	AC018939	45.8	12.4	134940	2	AC018939 Homo sapi
38	AC026508	45.8	12.4	154349	2	AC026508 Homo sapi
39	AC022783	45.8	12.4	166604	2	AC022783 Mus muscu
40	AC09640	45.8	12.4	178821	2	AC09640 Mus muscu
41	AC107846	45.6	12.3	51882	2	AC107846 Mus muscu
42	AC105985	45.6	12.3	72509	2	AC105985 Mus muscu
43	AC021584	45.6	12.3	78765	2	AC021584 Homo sapi
44	AC074204	45.6	12.3	152732	2	AC074204 Mus muscu
45	AC007610	45.6	12.3	230426	2	AC007610 Homo sapi
46	AC079429	45.6	12.3	258174	2	AC079429 Mus muscu

AC074220 89855 bp DNA linear HTG 25-JAN-2002
Mus musculus chromosome 17 clone RP23-76116 strain C57BL/6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

AC074220 7 GI:13876390
HTG: HTGS_PHASE1; HTGS_DRAFT.

house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 89855)

Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,

Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,

Gordon, M., Goltz, J.S. and Kucherlapati, R.

High Throughput Mouse Sequencing

Unpublished

2 (bases 1 to 89855)

Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,

Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,

Gordon, M., Goltz, J.S. and Kucherlapati, R.

Direct Submission

AC106591/c	AC106591	39811 bp	DNA	linear	HTG 12-JAN-2002
LOCUS	Rattus norvegicus clone CH230-212K18, *** SEQUENCING IN PROGRESS				
DEFINITION	***, 25 unordered pieces.				
ACCESSION	AC106591				
VERSION	AC106591.1	GI:18139115			
KEYWORDS	HTG; HTGS_PHASE1.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia; Eutheria: Rodentia: Sciurognathi; Muridae: Murinae; Rattus.				
AUTHORS	1 (bases 1 to 39811) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alzbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbieri,J., Benton,J., Blimie,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,N., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Joudah,S., Jacobson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., L., Louissegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., McInerney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,G., Nguyen,N., Nickerson,E., Nwokwoko,S., Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Slason,I., Sodergren,E., Sonaikar,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 39811)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT	Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GLLB Center clone name: CH230-212K18 ----- Summary Statistics Assembly program: Phrap; version 0.990329First call to findPhrapList Consensus quality: 38345 bases at least Q40 Consensus quality: 42938 bases at least Q30				
FEATURES	Source				
BASE COUNT	10663 a 8757 c 7160 g 10537 t 2694 others				
ORIGIN					
	1..39811 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="CH230-212K18"				
	3585: contig of 3585 bp in length 3685: gap of unknown length 3686: contig of 1612 bp in length 5297: gap of unknown length 5298: contig of 2129 bp in length 7526: gap of unknown length 7527: contig of 1690 bp in length 9316: gap of unknown length 9317: contig of 1831 bp in length 9417: gap of unknown length 11248: contig of 2193 bp in length 11348: gap of unknown length 13541: contig of 1439 bp in length 13640: gap of unknown length 15079: contig of 1490 bp in length 15179: gap of unknown length 15708: contig of 1354 bp in length 16668: gap of unknown length 16769: contig of 1484 bp in length 18123: gap of unknown length 18224: contig of 1332 bp in length 19707: gap of unknown length 19708: contig of 1179 bp in length 21139: gap of unknown length 21238: contig of 1513 bp in length 22418: gap of unknown length 22519: contig of 1145 bp in length 24032: gap of unknown length 24131: contig of 1065 bp in length 25277: gap of unknown length 25376: contig of 1633 bp in length 26442: gap of unknown length 26443: contig of 1094 bp in length 28174: gap of unknown length 28275: contig of 1007 bp in length 29369: gap of unknown length 30476: contig of 1487 bp in length 30575: gap of unknown length 30576: contig of 1447 bp in length 32063: gap of unknown length 33610: contig of 1538 bp in length 35248: gap of unknown length 35348: contig of 1019 bp in length 36367: gap of unknown length 36467: contig of 1011 bp in length 37478: gap of unknown length 37578: contig of 1032 bp in length 38610: gap of unknown length 38710: contig of 1102 bp in length.				

Query Match	13.7%; Score 50.6; DB 2; Length 39811;
Best Local Similarity	58.2%; Pred. No. 0.094;
Matches	89; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
yy	19 gcaagatgaggggcgccgcggggcgtagccaagccagggcgagggcgagggcgcg 78
bb	36123 GGTGGGTTGGGGCGCGGGCGCGGGCGGGTTCGCCTGGTGTTGGTGGCCGG 36064
yy	79 ggcctgtgggaagggtctggcgctggcaggacctggctggggtctcttggcccctgctg 138
bb	36083 GC CGCGGGGGCGCGGGCGGCCCTGGCGGGGGGTGGTGGGGGGCGCGGGGGCGCTGCTG 36004
yy	139 tttcttgtcggcaatgctgggtgctgtgactc 171
bb	36003 GTGTGCGGGCGCGCGCGCGGGCTCGGGGCTC 35971
RESULT	4
Locus	I66494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.
ACCESSION	I66494
VERSION	I66494.1 GI:2724471
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 7218)
AUTHORS	Dorner,F., Schelflinger,F. and Falkner,F.Gunter.
TITLE	Recombinant fowlpox virus
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;
FEATURES	Location/Qualifiers 1..7218 /organism="unknown"
BASE COUNT	1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN	
Query Match	13.6%; Score 50.4; DB 6; Length 7218;
Best Local Similarity	0.9%; Pred. No. 0.13;
Matches	3; Conservative 214; Mismatches 135; Indels 0; Gaps 0;
QY	6 ttgccacgaatacgaaagtgaaggcgccgcggcgccgtagccaacgcccgaggcg 65
Db	1440 TGTACRRR 1381
QY	66 cgagagggcgccgctgtggaaaggtctggcgcttcggcagacctggcggtgctc 125
Db	1380 RR 1321
QY	126 cttyggccctgtgtgttcggcaatctgggtgctgtgactcggataacctgga 185
Db	1320 RR 1261
QY	186 gatccctgttttggcgaaatccgggggttagtgtcatcaagaactagagtggtgg 245
Db	1260 RR 1201
QY	246 aggaaggtcttatcaggaagcctgtcgaaatgaagagtggccagggaaaacatgg 305
Db	1200 RR 1141
QY	306 cgtcagaggaactcactccgcacaaccacagagagcgagatgagga 357
Db	1140 RR 1089
RESULT	5
Locus	OCU05204/c 1396 bp mRNA linear MAM 27-JAN-1995
DEFINITION	Oryctolagus cuniculus preproacrosin mRNA, complete cds.
ACCESSION	U05204
VERSION	U05204.1 GI:451841

```

Db 969 GTGCCAGAGGGCGCTGGTGCACGCCGGGATCGGTAGTGGAGGGTAGGGGTAGC 910
Qy 259 tacagaagacctgctgcgaataagagtggtggccagggaaagcatggtgagcag 312
Db 909 TGGTTGAATCATGCGCAAGAGCATTAGAGCCGATTCTAGAACGATCCAGTCCAG 856

RESULT 6
AC023197/c
LOCUS AC023197
DEFINITION Mus musculus clone RP22-239F2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023197
VERSION AC023197.2 GI:7249227
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 141892)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Mus musculus chromosome, clone RP22-239F2
Unpublished
2 (bases 1 to 141892)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bida, F.,
Boguslavskiy, L., Boukhgaiter, B., Brown, A., Burkett, G., Castle, A.,
Chopel, Y., Collangelo, M., Collins, S., Collimore, A., Cooke, P.,
DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A., and Zody, M.
Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6957784.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5917
Center clone name: 239_F_2
-----
* NOTE: This record contains 155 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 656: contig of 656 bp in length
* 657 756: gap of 100 bp
* 757 1510: contig of 754 bp in length
* 1511 1610: gap of 100 bp
* 1611 2384: contig of 754 bp in length
*
2365 2464: gap of 100 bp
2465 3234: contig of 770 bp in length
3235 3334: gap of 100 bp
3335 4083: contig of 749 bp in length
4084 4183: gap of 100 bp
4184 4931: contig of 748 bp in length
4932 5031: gap of 100 bp
5032 5787: contig of 756 bp in length
5788 5887: gap of 100 bp
5888 6644: contig of 757 bp in length
6645 6744: gap of 100 bp
6745 7516: contig of 772 bp in length
7517 7616: gap of 100 bp
7617 8376: contig of 760 bp in length
8377 8476: gap of 100 bp
8477 9227: contig of 751 bp in length
9228 9327: gap of 100 bp
9328 10089: contig of 762 bp in length
10090 10189: gap of 100 bp
10190 10952: contig of 763 bp in length
10953 11052: gap of 100 bp
11053 11826: contig of 774 bp in length
11827 11925: gap of 100 bp
11927 12688: contig of 762 bp in length
12689 12788: gap of 100 bp
12789 13543: contig of 755 bp in length
13544 13643: gap of 100 bp
13644 14407: contig of 764 bp in length
14408 14507: gap of 100 bp
14508 15275: contig of 768 bp in length
15276 15375: gap of 100 bp
15376 16144: contig of 769 bp in length
16145 16244: gap of 100 bp
16245 17002: contig of 758 bp in length
17003 17102: gap of 100 bp
17103 17803: contig of 701 bp in length
17804 17903: gap of 100 bp
17904 18618: contig of 715 bp in length
18619 18718: gap of 100 bp
18719 19486: contig of 768 bp in length
19487 19586: gap of 100 bp
19587 20356: contig of 770 bp in length
20357 20456: gap of 100 bp
20457 21222: contig of 766 bp in length
21223 21322: gap of 100 bp
21323 22069: contig of 747 bp in length
22070 22169: gap of 100 bp
22170 22921: contig of 752 bp in length
22922 23021: gap of 100 bp
23022 23772: contig of 751 bp in length
23773 23872: gap of 100 bp
23873 24634: contig of 762 bp in length
24635 24734: gap of 100 bp
24735 25363: contig of 629 bp in length
25364 25463: gap of 100 bp
25464 26227: contig of 764 bp in length
26228 26327: gap of 100 bp
26328 27088: contig of 761 bp in length
27089 27188: gap of 100 bp
27189 27955: contig of 767 bp in length
27956 28055: gap of 100 bp
28056 28836: contig of 781 bp in length
28837 28936: gap of 100 bp
28937 29702: contig of 766 bp in length
29703 29802: gap of 100 bp
29803 30577: contig of 775 bp in length
30578 30677: gap of 100 bp
30678 31482: contig of 805 bp in length
31483 31582: gap of 100 bp
31583 32341: contig of 759 bp in length
32342 32441: gap of 100 bp
32442 33202: contig of 761 bp in length
33203 33302: gap of 100 bp

```


Query Match	13.2%	Score 48.8	DB 2	Length 236301		
Best Local Similarity	50.4%	Pred. No. 0.2				
Matches 119	Conservative	0	Mismatches 117	Indels 0	Gaps 0	
QY	18	tgcgaaggtgagggcgccggccgagggcgtagccagcccgagggcgagggcgagggcg	77			
DB	89602	TGGAGCGGGGGGGGGGGGGGGGGAGTGGGGGTGGAGGGGGGGGGGGGGGGAGGGA	89661			
QY	78	gggcctgtggaaggtctggccctggcagggacccctgggctggtggtctctcttggccctgct	137			
DB	89662	GGGGGTCCGGAACGAGGAGGGGGGGGGGGATGGGGGGGGGGGGGGGGGGGGGGGT	89721			
QY	138	gtgtgcttgcggcaatgtggtgtgtgactctcgataaacctggagatccctgcttt	197			
DB	89722	GTGTGCGGGGTGGGGGGGGGGGGAGGTGGGTGGGGGGGGGGGGGGGGTGGGGGG	89781			
QY	198	tggcggaatccgggtagttgtctcatcaagactagatggtgggtggagggaagg	253			
DB	89782	AGGGGGGGGGGGGGGGGGGTGGGGAGGAGGGGGGTGGGGGGGGGGGGGGGGGG	89837			
RESULT	8					
AC095701/c		163525 bp	DNA	linear	HTG 20-DEC-2001	
LOCUS		Rattus norvegicus clone CH230-9B18, ***	SEQUENCING IN PROGRESS ***			
DEFINITION		62 unordered pieces.				
ACCESSION		AC095701.2	GI:17942281			
VERSION		HTG; HTGS_PHASE1.				
KEYWORDS		Norway rat.				
SOURCE		Rattus norvegicus				
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE		1 (bases 1 to 163525)				
AUTHORS		Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.I., Amaratunga,H.C., Aze,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.I., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Denny-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nickerson,E., Nwankwo,S., Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.B., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,				

Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
 2 (bases 1 to 148921)
 Worley, K.C.
 Direct Submission
 Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15625779.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GDKJ
 Center clone name: CH230-9F14
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329first call to
 findphraplist

Consensus quality: 117844 bases at least Q40
 Consensus quality: 130606 bases at least Q30
 Consensus quality: 138533 bases at least Q20
 Estimated insert size: 78668; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 87 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3865: contig of 3865 bp in length
 3866 3955: gap of unknown length
 3966 6975: contig of 3010 bp in length
 6976 7075: gap of unknown length
 7076 10377: contig of 3302 bp in length
 10378 10478: gap of unknown length
 10479 13677: contig of 3200 bp in length
 13678 13777: gap of unknown length
 13778 17257: contig of 3480 bp in length
 17258 17357: gap of unknown length
 17358 18597: contig of 1240 bp in length
 18598 18697: gap of unknown length
 18698 20220: contig of 1523 bp in length
 20221 20320: gap of unknown length
 20321 23582: contig of 3262 bp in length
 23583 23682: gap of unknown length
 23683 25938: contig of 2256 bp in length
 25939 26038: gap of unknown length
 26039 28627: contig of 2589 bp in length
 28628 28727: gap of unknown length
 28728 29791: contig of 1064 bp in length
 29792 29891: gap of unknown length
 29892 31945: contig of 2054 bp in length
 31946 32045: gap of unknown length
 32046 35046: contig of 3009 bp in length
 35047 35154: gap of unknown length
 35155 37831: contig of 2677 bp in length
 37832 37931: gap of unknown length
 40056: contig of 2125 bp in length
 40057 40156: gap of unknown length
 40157 41954: contig of 1798 bp in length
 41955 42054: gap of unknown length
 42055 43452: contig of 1398 bp in length

43453 43552: gap of unknown length
 43553 45687: contig of 2135 bp in length
 45688 45787: gap of unknown length
 45788 47477: contig of 1690 bp in length
 47478 47577: gap of unknown length
 47578 49684: contig of 2107 bp in length
 49685 49784: gap of unknown length
 49785 51965: contig of 2181 bp in length
 51966 52065: gap of unknown length
 52066 53630: contig of 1565 bp in length
 53631 54995: gap of unknown length
 54996 57371: contig of 1266 bp in length
 57372 58096: gap of unknown length
 58097 58160: contig of 1064 bp in length
 58161 58260: gap of unknown length
 58261 58386: contig of 1126 bp in length
 58387 58486: gap of unknown length
 58487 58795: contig of 1309 bp in length
 58796 58895: gap of unknown length
 58896 60547: contig of 1652 bp in length
 60548 60647: gap of unknown length
 60649 63005: contig of 2358 bp in length
 63006 63105: gap of unknown length
 63106 65100: contig of 1995 bp in length
 65101 65200: gap of unknown length
 65201 66722: contig of 1522 bp in length
 66723 66822: gap of unknown length
 66823 69735: contig of 2913 bp in length
 69736 69835: gap of unknown length
 69836 71274: contig of 1439 bp in length
 71275 71374: gap of unknown length
 71375 72856: contig of 1482 bp in length
 72857 72956: gap of unknown length
 72957 74011: contig of 1055 bp in length
 74012 74111: gap of unknown length
 74112 75778: contig of 1667 bp in length
 75779 75878: gap of unknown length
 75879 77264: contig of 1386 bp in length
 77265 77364: gap of unknown length
 77365 79442: contig of 2278 bp in length
 79443 79643: gap of unknown length
 79644 81472: contig of 1730 bp in length
 81473 81572: gap of unknown length
 81573 83102: contig of 1530 bp in length
 83103 83202: gap of unknown length
 83203 84298: contig of 1096 bp in length
 84299 84398: gap of unknown length
 84399 85406: contig of 1008 bp in length
 85407 85506: gap of unknown length
 85507 86520: contig of 1014 bp in length
 86521 86620: gap of unknown length
 86621 87692: contig of 1072 bp in length
 87693 87792: gap of unknown length
 87793 89176: contig of 1384 bp in length
 89177 89276: gap of unknown length
 89277 90575: contig of 1299 bp in length
 90576 91943: gap of unknown length
 91944 92043: gap of unknown length
 92044 93415: contig of 1372 bp in length
 93416 93515: gap of unknown length
 93516 94544: contig of 1029 bp in length
 94545 94644: gap of unknown length
 94645 95993: contig of 1349 bp in length
 95994 96093: gap of unknown length
 96094 97263: contig of 1070 bp in length
 97264 97263: gap of unknown length
 97264 99162: contig of 1899 bp in length
 99163 99262: gap of unknown length
 99263 100319: contig of 1057 bp in length
 100320 100419: gap of unknown length
 100420 102042: contig of 1623 bp in length
 102043 102142: gap of unknown length


```

* 28187 29106: contig of 920 bp in length
* 29107 29206: gap of 100 bp
* 29207 30153: contig of 949 bp in length
* 30156 30255: gap of 100 bp
* 30256 31182: contig of 927 bp in length
* 31183 31282: gap of 100 bp
* 31283 32108: contig of 826 bp in length
* 32109 32208: gap of 100 bp
* 32209 33133: contig of 925 bp in length
* 33134 33233: gap of 100 bp
* 33234 34221: contig of 988 bp in length
* 34222 34321: gap of 100 bp
* 34322 35205: contig of 884 bp in length
* 35206 35305: gap of 100 bp
* 35306 36272: contig of 967 bp in length
* 36273 36372: gap of 100 bp
* 36373 37360: contig of 988 bp in length
* 37361 37460: gap of 100 bp
* 37461 38362: contig of 902 bp in length
* 38363 38462: gap of 100 bp
* 38463 39425: contig of 963 bp in length
* 39426 39525: gap of 100 bp
* 39526 40391: contig of 866 bp in length
* 40392 40491: gap of 100 bp
* 40492 41461: contig of 970 bp in length
* 41462 42548: contig of 987 bp in length
* 42549 42648: gap of 100 bp
* 42649 43642: contig of 994 bp in length
* 43643 43742: gap of 100 bp
* 43743 44701: contig of 959 bp in length
* 44702 44801: gap of 100 bp
* 44802 45709: contig of 908 bp in length
* 45710 45809: gap of 100 bp
* 45810 46742: contig of 933 bp in length.
FEATURES
    source
        1. .46742
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-94114"
            /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT 11493 a 8700 c 8777 g 12424 t 5348 others
ORIGIN

Query Match      12.98; Score 47.6; DB 2; Length 46742;
Best Local Similarity 30.4%; Pred. No. 0.44;
Matches 92; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 13 gcaaatgcgaaggtagggggcgccgcggggcgtagccaaagcccgagggcgagg 72
Db 12752 GNNNGNNNNNGNGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 12811

QY 73 gggcgggcgctgtggaagggctggtggcctgagcagacctggctgggtctcttggcc 132
Db 12812 NGNGGGGGGGGNNNGNNNGNNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNN 12871

QY 133 ctgctgtgtcttgcgcaatgctgggtgctgtgactctcgataacatggagatccct 192
Db 12872 NGNNNGGGGGGNNNGNNNGGNGGNGGNGGNNNGNNNGNNNGNNNGNNNNNNNN 12931

QY 193 gcttttggcgaatccgggggtagttgtctatcaagactagagtggtggggtgaggaag 252
Db 12932 GGNNNNNNGGNGGNGGNGGNGGNGGNNNNNNNNNGGNGGNGGNGGNGGNGG 12991

QY 253 gcttcaacagaagctctcgaatgaagattggcaggggaagcatggcgtgcag 312
Db 12992 GNNNGGNGGGGGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 13051

QY 313 agg 315
Db 13052 GGG 13054
```

```

AC093968      199536 bp      DNA      linear      HTG 20-DEC-2001
Rattus norvegicus clone CH230-6J18, *** SEQUENCING IN PROGRESS ***
57 unordered pieces.
AC093968      AC093968.4 GI:17941700
HTG: HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 199536)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burket,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenkwo,S.,
Ogoh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Roife,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 199536)
Worley,K.C.
Direct Submission
Submitted (13-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062167.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc.help@bcm.tmc.edu
----- Project Information
Center project name: GBRK
Center clone name: CH230-6J18
----- Summary Statistics
-----
Assembly program: Phrap; version 0.990329First call to
```


Fri May 24 10:49:50 2002

Qy 185 agatccctgttttggcggaatccggggtagttgtctcatcaagactagaggtgggggtg 244
 Db 197572 TCACGCGCGCGTGGCGGAACCGGTGCGTACACGCGGGAAGCGTGTGGTGCCGCC 197513

Qy 245 gaggaa 251
 Db 197512 GGGGTA 197506

RESULT 12
 AC095175/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-8J16, *** SEQUENCING IN PROGRESS ***
 60 unordered pieces.
 AC095175
 HTG: HTGS-PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 107781)
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, F.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsif, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudas, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, D.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogah, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vazquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 107781)
 Worley, K.C.
 Direct Submission
 Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15625729.
 ----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GC2P
 Center clone name: CH230-8J16
 Summary Statistics
 Assembly program: Phrap; version 0.990329First call to findPhrapList
 Consensus quality: 73837 bases at least Q40
 Consensus quality: 87577 bases at least Q30
 Consensus quality: 95671 bases at least Q20
 Estimated insert size: 65224; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently consists of 60 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
 1 3529: contig of 3529 bp in length
 3530 3629: gap of unknown length
 3630 7406: contig of 3777 bp in length
 7407 7506: gap of unknown length
 7507 10229: contig of 2723 bp in length
 10230 10329: gap of unknown length
 10330 12998: contig of 2669 bp in length
 12999 13098: gap of unknown length
 13099 14993: contig of 1795 bp in length
 14994 14993: gap of unknown length
 14994 17445: contig of 2552 bp in length
 17445 17445: gap of unknown length
 17446 19280: contig of 1635 bp in length
 19281 19380: gap of unknown length
 21537 21636: contig of 2156 bp in length
 21636 23581: gap of unknown length
 23581 23581: gap of unknown length
 23582 25999: contig of 2118 bp in length
 25999 25999: gap of unknown length
 25999 28806: contig of 2907 bp in length
 28806 28806: gap of unknown length
 28807 30583: contig of 1677 bp in length
 30583 30583: gap of unknown length
 30584 32640: contig of 1957 bp in length
 32640 32640: gap of unknown length
 32641 34130: contig of 1390 bp in length
 34130 34230: gap of unknown length
 34230 35751: contig of 1521 bp in length
 35751 35751: gap of unknown length
 35752 37605: contig of 1754 bp in length
 37605 37605: gap of unknown length
 37606 40764: contig of 3059 bp in length
 40764 40864: gap of unknown length
 40864 43662: contig of 2798 bp in length
 43662 43662: gap of unknown length
 43663 44948: contig of 1186 bp in length
 44948 45048: gap of unknown length
 45048 46381: contig of 1333 bp in length
 46381 46381: gap of unknown length
 46382 47524: contig of 1043 bp in length
 47524 47524: gap of unknown length
 47525 48993: contig of 1375 bp in length
 48993 49000: gap of unknown length
 49000 50960: contig of 1861 bp in length
 50960 51060: gap of unknown length

Db	90874	GGGAGCGCGGGAGAGCGCGCGCGGTGGGGAGCGCGGGGGCGCGGGGGGGCGG	90815
Qy	79	ggcctgtgggaaggtctcggccctggcaggacctgggctctgggctctcttggccctactg	138
Db	90814	CGCGGAGGAGCGGGCGGGCGGATGGGAGGAGCGGGGGGGGGGCGGTGACCTCGTG	90755
Qy	139	tgtcctttggcgaatcctcgtgggtgctgtgactctcg	174
Db	90754	GGCGGAGGAGCGTATCGCGGGCGGGTACCGCG	90719
RESULT	13		
AC079431			
LOCUS			
DEFINITION		AC079431 300695 bp DNA linear HTG 01-SEP-2000	
ACCESSION		77 unordered pieces.	
VERSION		AC079431	
KEYWORDS		HTG: HTGS_PHASE1; HTGS_DRAFT.	
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
TITLE		DOE Joint Genome Institute.	
JOURNAL		Sequencing of Mouse	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 300695)	
TITLE		DOE Joint Genome Institute.	
JOURNAL		Direct Submission	
COMMENT		Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA -----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: http://www.jgi.doe.gov ----- Project Information Project Name: 0 Center clone name: RPCI-23_LJ12 ----- Summary Statistics Consensus quality: 164936 bases at least Q40 Consensus quality: 197223 bases at least Q30 Consensus quality: 214345 bases at least Q20 Estimated insert size: 257300; agarose-fp estimation Quality coverage: 3.01 in Q20 bases; agarose-fp estimation Quality coverage: 2.65 in Q20 bases; sum-of-contigs estimation. * NOTE: This is a 'working draft' sequence. It currently * consists of 77 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 1076: contig of 1076 bp in length * 1077 1176: gap of unknown length * 1177 2208: contig of 1032 bp in length * 2308 2309: gap of unknown length * 3527 3528: contig of 1219 bp in length * 3627 3628: gap of unknown length * 4764 4765: contig of 1137 bp in length * 4864 4865: gap of unknown length * 5952 5953: contig of 1088 bp in length * 6052 6053: gap of unknown length * 7182 7183: contig of 1130 bp in length * 7282 7283: gap of unknown length * 8420 8421: contig of 1138 bp in length * 8520 8521: gap of unknown length * 9616 9617: contig of 1096 bp in length * 9716: gap of unknown length	

Fri May 24 10:49:50 2002

* 9717 11129: contig of 1413 bp in length
* 11130 11229: gap of unknown length
* 11230 12408: contig of 1179 bp in length
* 12409 12508: gap of unknown length
* 12509 13767: contig of 1259 bp in length
* 13768 13867: gap of unknown length
* 13868 15044: contig of 1177 bp in length
* 15045 15144: gap of unknown length
* 15145 16325: contig of 1181 bp in length
* 16326 16425: gap of unknown length
* 16426 17578: contig of 1153 bp in length
* 17579 17678: gap of unknown length
* 17679 18845: contig of 1167 bp in length
* 18846 18945: gap of unknown length
* 18946 20060: contig of 1115 bp in length
* 20061 20160: gap of unknown length
* 20161 21258: contig of 1098 bp in length
* 21259 21358: gap of unknown length
* 21359 22390: contig of 1032 bp in length
* 22391 22490: gap of unknown length
* 22491 23599: contig of 1109 bp in length
* 23599 23699: gap of unknown length
* 23700 24907: contig of 1208 bp in length
* 24908 25007: gap of unknown length
* 25008 26121: contig of 1114 bp in length
* 26122 26221: gap of unknown length
* 26222 27273: contig of 1052 bp in length
* 27273 28472: gap of unknown length
* 28472 28572: gap of unknown length
* 28573 29683: contig of 1111 bp in length
* 29684 29783: gap of unknown length
* 29784 30834: contig of 1051 bp in length
* 30835 30934: gap of unknown length
* 30935 32135: contig of 1101 bp in length
* 32136 33334: contig of 1199 bp in length
* 33335 33434: gap of unknown length
* 33435 34598: contig of 1164 bp in length
* 34599 34698: gap of unknown length
* 34699 35818: contig of 1120 bp in length
* 35819 35918: gap of unknown length
* 35919 37082: contig of 1164 bp in length
* 37083 37182: gap of unknown length
* 37183 38645: contig of 1463 bp in length
* 38646 38745: gap of unknown length
* 38746 39768: contig of 1023 bp in length
* 39769 39868: gap of unknown length
* 39869 40973: contig of 1105 bp in length
* 40974 41073: gap of unknown length
* 41074 42174: contig of 1101 bp in length
* 42175 42274: gap of unknown length
* 42275 43341: contig of 1067 bp in length
* 43342 43441: gap of unknown length
* 43442 44532: contig of 1091 bp in length
* 44533 44632: gap of unknown length
* 44633 46068: contig of 1436 bp in length
* 46069 46168: gap of unknown length
* 46169 47421: contig of 1253 bp in length
* 47422 47521: gap of unknown length
* 47522 48629: contig of 1108 bp in length
* 48630 48729: gap of unknown length
* 48730 49986: contig of 1257 bp in length
* 49987 50086: gap of unknown length
* 50087 51394: contig of 1308 bp in length
* 51395 51494: gap of unknown length
* 51495 53158: contig of 1664 bp in length
* 53159 53258: gap of unknown length
* 53259 54802: contig of 1544 bp in length
* 54803 54902: gap of unknown length
* 54903 56009: contig of 1107 bp in length
* 56010 56109: gap of unknown length
* 56110 57452: contig of 1343 bp in length

* 57453 57552: gap of unknown length
* 57553 58667: contig of 1115 bp in length
* 58668 58767: gap of unknown length
* 58768 59924: contig of 1157 bp in length
* 59925 60024: gap of unknown length
* 60025 61159: contig of 1135 bp in length
* 61160 61259: gap of unknown length
* 61260 62764: contig of 1505 bp in length
* 62765 62864: gap of unknown length
* 62865 64293: contig of 1429 bp in length
* 64294 64393: gap of unknown length
* 64394 65793: contig of 1400 bp in length
* 65794 65893: gap of unknown length
* 65894 67168: contig of 1275 bp in length
* 67169 67268: gap of unknown length
* 67269 69245: gap of unknown length
* 69246 69345: gap of unknown length
* 69346 70850: contig of 1505 bp in length
* 70851 70950: gap of unknown length
* 70951 73220: contig of 2270 bp in length
* 73221 73320: gap of unknown length
* 73321 76908: contig of 3588 bp in length
* 76909 77008: gap of unknown length
* 77009 78773: contig of 1765 bp in length
* 78774 78873: gap of unknown length
* 78874 81789: contig of 2916 bp in length
* 81790 81889: gap of unknown length
* 81890 84399: contig of 2510 bp in length
* 84400 84499: gap of unknown length
* 84500 87255: gap of unknown length
* 87256 87356: gap of unknown length
* 87357 91360: contig of 4005 bp in length
* 91361 91460: gap of unknown length
* 91461 95576: contig of 4116 bp in length
* 95577 95676: gap of unknown length
* 95677 99941: contig of 4265 bp in length
* 99942 100041: gap of unknown length
* 100042 105431: contig of 5390 bp in length
* 105432 105531: gap of unknown length
* 105532 112490: contig of 6959 bp in length
* 112491 112591: gap of unknown length
* 112591 121967: contig of 9377 bp in length
* 121968 122067: gap of unknown length
* 122068 132042: contig of 9975 bp in length
* 132043 132142: gap of unknown length
* 132143 14368: contig of 14226 bp in length
* 14369 146468: gap of unknown length
* 146469 156606: contig of 10138 bp in length
* 156607 156706: gap of unknown length
* 156707 166814: contig of 10108 bp in length
* 166815 16914: gap of unknown length
* 16915 177819: contig of 10905 bp in length
* 177820 177919: gap of unknown length
* 177920 191863: contig of 13944 bp in length
* 191864 191963: gap of unknown length
* 191964 204880: contig of 12917 bp in length
* 204881 204980: gap of unknown length
* 204981 220059: contig of 15079 bp in length
* 220060 220159: gap of unknown length
* 220160 238639: contig of 18480 bp in length
* 238640 238739: gap of unknown length
* 238740 262566: contig of 23827 bp in length
* 262567 262666: gap of unknown length
* 262667 300695: contig of 38029 bp in length.

FEATURES

source

Location/Qualifiers
1. .300695
/organism="Mus musculus"

Query Match 12.8%; Score 47.2; DB 2; Length 300695;
Best Local Similarity 42.0%; Pred. No. 0.44;
Matches 97; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 19 gcaagagtggggggcgccgcggcgtagcccaagcccgaggggcgagggcgcg 78

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 192182)
Worley,K.C.
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624432.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAZD
Center clone name: CH230-4A13
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 135751 bases at least Q40
Consensus quality: 146303 bases at least Q30
Consensus quality: 153879 bases at least Q20
Estimated insert size: 123383; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 94 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 5533: contig of 5533 bp in length
* 5534: gap of unknown length
* 5634: contig of 6404 bp in length
* 12038: gap of unknown length
* 12137: gap of unknown length
* 12138: contig of 4982 bp in length
* 17120: gap of unknown length
* 22077: contig of 4858 bp in length
* 22078: gap of unknown length
* 22178: contig of 3062 bp in length
* 25230: gap of unknown length
* 25240: contig of 4911 bp in length
* 25340: gap of unknown length
* 30251: contig of 3461 bp in length
* 30351: gap of unknown length
* 33812: contig of 3461 bp in length
* 33911: gap of unknown length
* 37447: contig of 3535 bp in length
* 37448: gap of unknown length
* 37547: contig of 2816 bp in length
* 40363: gap of unknown length
* 40463: contig of 3075 bp in length
* 43537: gap of unknown length
* 43638: contig of 3241 bp in length
* 46878: gap of unknown length
* 46979: contig of 3376 bp in length
* 50354: gap of unknown length
* 50454: contig of 2235 bp in length
* 52689: gap of unknown length
* 52690: contig of 3642 bp in length
* 52790: gap of unknown length
* 56431: contig of 2675 bp in length
* 56531: gap of unknown length
* 56532: contig of 2200 bp in length
* 59207: gap of unknown length
* 59307: contig of 2200 bp in length
* 61507: gap of unknown length
* 61607: contig of 3075 bp in length
* 64682: gap of unknown length
* 64781: contig of 2423 bp in length
* 67205: gap of unknown length
* 67305: contig of 1551 bp in length
* 68855: gap of unknown length
* 68956: contig of 3485 bp in length
* 72441: gap of unknown length
* 72541: contig of 1787 bp in length
* 74328: gap of unknown length
* 74428: contig of 3037 bp in length
* 77465: gap of unknown length
* 77564: contig of 1703 bp in length
* 79267: gap of unknown length
* 79268: contig of 2184 bp in length
* 79368: gap of unknown length
* 81552: contig of 1519 bp in length
* 81652: gap of unknown length
* 83171: contig of 1479 bp in length
* 83271: gap of unknown length
* 84750: contig of 1333 bp in length
* 84849: gap of unknown length
* 86439: contig of 1875 bp in length
* 86439: gap of unknown length
* 86439: contig of 1875 bp in length
* 86539: gap of unknown length
* 88414: contig of 1760 bp in length
* 88514: gap of unknown length
* 90274: contig of 2080 bp in length
* 90374: gap of unknown length
* 92454: contig of 2779 bp in length
* 92554: gap of unknown length
* 95333: contig of 1940 bp in length
* 95433: gap of unknown length
* 96765: contig of 1333 bp in length
* 96865: gap of unknown length
* 96866: contig of 2433 bp in length
* 99299: gap of unknown length
* 99399: contig of 1534 bp in length
* 100933: gap of unknown length
* 101033: contig of 1573 bp in length
* 101333: gap of unknown length
* 102606: contig of 1573 bp in length
* 102706: gap of unknown length
* 104646: contig of 1940 bp in length
* 104746: gap of unknown length
* 106521: contig of 1775 bp in length
* 106521: gap of unknown length
* 106621: contig of 1267 bp in length
* 107888: gap of unknown length
* 107988: contig of 1107 bp in length
* 109095: gap of unknown length
* 109195: contig of 2007 bp in length
* 111202: gap of unknown length
* 111302: contig of 2182 bp in length
* 113484: gap of unknown length
* 113584: contig of 1513 bp in length
* 115097: gap of unknown length
* 115197: contig of 2118 bp in length
* 117315: gap of unknown length
* 117415: contig of 1222 bp in length
* 118637: gap of unknown length
* 118737: contig of 2296 bp in length
* 121033: gap of unknown length
* 121133: contig of 1445 bp in length
* 122578: gap of unknown length
* 122678: contig of 2394 bp in length
* 125072: gap of unknown length
* 125172: contig of 1940 bp in length
* 127112: gap of unknown length
* 127212: contig of 1223 bp in length
* 128435: gap of unknown length
* 128535: contig of 1454 bp in length
* 129988: gap of unknown length
* 130088: contig of 1014 bp in length
* 130089: gap of unknown length
* 131103: contig of 1726 bp in length
* 131203: gap of unknown length
* 132928: contig of 1726 bp in length
* 133028: gap of unknown length

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

* 133029 134231: contig of 1203 bp in length
* 134232 134331: gap of unknown length
* 134332 135816: contig of 1485 bp in length
* 135817 135916: gap of unknown length
* 135917 137194: contig of 1278 bp in length
* 137195 137294: gap of unknown length
* 137295 139200: contig of 1906 bp in length
* 139201 139300: gap of unknown length

Query Match 12.7%; Score 47; DB 2; Length 192182;
Best Local Similarity 46.9%; Pred. No. 0.51;
Matches 113; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 13 gcaaatcgcaagtgagggcgccgagggcgtagccaagcccgagggcgaggagg 72
Db 16776 GGAAGGA 16717

QY 73 gggcgggcctgtggaagggtctggcctggcagaccctgggctggggtctccttgccc 132
Db 16716 GGG 16657

QY 133 ctgctgtgtgttgaggcaatgctgggtgtgtgactctcgataaacctgagatccct 192
Db 16656 GGG 16597

QY 193 gcttttgggcgaatccgggggtagtgtctcatcaagactagaggtggggtgaggggaag 252
Db 16596 GNNNGGG 16537

QY 253 g 253
Db 16536 G 16536

Search completed: May 24, 2002, 00:23:51
Job time: 11937 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 21:04:54 ; Search time 3953.4 Seconds
(without alignments)
132.333 Million cell updates/sec

Title: US-09-696-686-45

Perfect score: 25

Sequence: 1 agctcagacatggactccatgcc 25

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_hgt_hum.*

31: em_hgt_inv.*

32: em_hgt_other.*

33: em_hgtg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

c	1	25	100.0	1882	10	BC013549
c	2	25	100.0	1909	10	AF361644
c	3	25	100.0	89855	2	AC074220
c	4	20.8	83.2	187142	2	AC097589
c	5	19.8	79.2	109963	2	AC107207
c	6	19.8	79.2	184288	2	AC011401
c	7	19.8	79.2	186780	2	AC005740
c	8	19.2	76.8	962	9	HSU80744
c	9	19.2	76.8	1422	9	BC008898
c	10	19.2	76.8	1422	9	BC008961
c	11	19.2	76.8	1720	9	BC004423
c	12	19.2	76.8	18120	9	HSB20F6
c	13	19.2	76.8	113109	9	HS475N16
c	14	19.2	76.8	171247	9	AC007881
c	15	19.2	76.8	191719	2	AL671848
c	16	19.2	76.8	227238	2	AL669855
c	17	18.8	75.2	105686	9	AC012072
c	18	18.8	75.2	129947	9	AL355518
c	19	18.8	75.2	170796	2	AC069416
c	20	18.8	75.2	182268	2	AC015591
c	21	18.6	74.4	643	9	HS1TAP1
c	22	18.6	74.4	643	9	HS2TAP1
c	23	18.6	74.4	1446	9	AF117647
c	24	18.6	74.4	1575	9	AB028645
c	25	18.6	74.4	1584	9	AF117646
c	26	18.6	74.4	3042	9	HS4430383
c	27	18.6	74.4	3461	10	MMU430384
c	28	18.6	74.4	9662	9	PTR251484
c	29	18.6	74.4	9672	9	GGO251485
c	30	18.6	74.4	11791	1	AE004763
c	31	18.6	74.4	27278	2	AC014009
c	32	18.6	74.4	32941	1	SCM10
c	33	18.6	74.4	37472	9	AC004539
c	34	18.6	74.4	37892	2	AC016540
c	35	18.6	74.4	66109	6	AX330351
c	36	18.6	74.4	66109	9	HSNHCAPG
c	37	18.6	74.4	80765	8	NCB1308
c	38	18.6	74.4	102677	2	AC095370
c	39	18.6	74.4	106093	2	AC106270
c	40	18.6	74.4	131906	2	AC083983
c	41	18.6	74.4	142215	2	AC092066
c	42	18.6	74.4	160757	2	AC087513
c	43	18.6	74.4	163183	3	AC009842
c	44	18.6	74.4	171979	2	AC092516
c	45	18.6	74.4	173682	2	AC021316

ALIGNMENTS

RESULT 1

BC013549/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC013549 1882 bp mRNA linear ROD 07-SEP-2001
Mus musculus, RIKEN cDNA 1600025D17 gene, clone MGC:19340
IMAGE:4222133, mRNA, complete cds.

BC013549

BC013549.1 GI:15488835

MGC.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 1882)

Strausberg, R.

Direct Submission

Submitted (04-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) md@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 25 Row: j Column: 9
 Location/Qualifiers

FEATURES
 source 1..1882
 /organism="Mus musculus"
 /db_xref="locusID:72029"
 /db_xref="taxon:10090"
 /clone="MGC:19340 IMAGE:4222133"
 /tissue_type="Kidney, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 91..921
 /codon_start=1
 /product="RIKEN cDNA 1600025D17 gene"
 /protein_id="AAH13549.1"
 /db_xref="GI:15488836"
 /translation="MESMSELAPRCLLPFLPILLPLPAPKLGSPAGAEETDWR
 LPSKEVCYKAVVELKSAFETGKTKEVIDTGILGDKSGVKYTKSDRLRIETET
 ICKRLDYSLHRTGSRNFAKMSFTFETLNLVHKGKVVYMDIPYELNWTSAEVA
 DLKQCDVLVEFEFIEDWYRNHOEDLTFELCANHLVKGKDTSCLAERWSGKKGDI
 ASLGGKSKKRRSGVKGSSGSSGSKORKEGLGLEDANAEIEEGVQKASPLPHSPDPDL
 "

BASE COUNT 432 a 539 c 549 g 362 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 10; Length 1882;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agctcagacatgactccatgccc 25
 |||||
 Db 110 AGCTCAGACATGGACTCCATGCCCC 86

RESULT 2
 AF361644/c 1909 bp mRNA linear ROD 12-MAY-2001
 LOCUS Mus musculus putative retinoic acid-regulated protein mRNA,
 complete cds.
 DEFINITION
 AF361644
 VERSION AF361644.1 GI:14028766
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1909)
 REFERENCE Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
 Trapping and characterization of novel retinoic acid response
 elements
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1909)
 AUTHORS Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAR-2001) Biology, University of South Florida, 4202
 E. Fowler Avenue, Tampa, FL 33620, USA
 Location/Qualifiers

FEATURES
 source 1..1909
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="P19; derived from matings between C3H/He

females and male animals carrying an X chromosome derived from a feral mouse (McBurney & Rogers, Dev. Biol., 89, 503)"
 /cell_type="embryonal carcinoma"
 /clone_lib="24 hr retinoic acid-treated P19 cDNA (Bouillet et al., Dev. Biol., 170, 420)"
 89..919
 CDS

/codon_start=1
 /product="putative retinoic acid-regulated protein"
 /protein_id="AAK52494.1"
 /db_xref="GI:14028767"
 /translation="MESMSELAPRCLLPFLPILLPLPAPKLGSPAGAEETDWR
 LPSKEVCYKAVVELKSAFETGKTKEVIDTGILGDKSGVKYTKSDRLRIETET
 ICKRLDYSLHRTGSRNFAKMSFTFETLNLVHKGKVVYMDIPYELNWTSAEVA
 DLKQCDVLVEFEFIEDWYRNHOEDLTFELCANHLVKGKDTSCLAERWSGKKGDI
 ASLGGKSKKRRSGVKGSSGSSGSKORKEGLGLEDANAEIEEGVQKASPLPHSPDPDL
 "

BASE COUNT 424 a 536 c 558 g 386 t 5 others
 ORIGIN
 Query Match 100.0%; Score 25; DB 10; Length 1909;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agctcagacatgactccatgccc 25
 |||||
 Db 108 AGCTCAGACATGGACTCCATGCCCC 84

RESULT 3
 AC074220 89855 bp DNA linear HTG 25-JAN-2002
 LOCUS Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
 WORKING DRAFT SEQUENCE, 38 unordered pieces.
 DEFINITION
 AC074220
 VERSION AC074220.7 GI:13876390
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 89855)
 AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
 High Throughput Mouse Sequencing
 Unpublished
 2 (bases 1 to 89855)
 REFERENCE Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
 Direct Submission
 JOURNAL Submitted (20-JUL-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

COMMENT
 On Apr 28, 2001 this sequence version replaced gi:11276104.
 -----Genome Center
 Center: Harvard Partners Genome Center
 Center Code: HPGC
 Web site: <http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mouse.html>
 Contact: gnkrm@capecod.bwh.harvard.edu
 e.html
 -----Summary Statistics
 Center project name: ADY
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 *Consensus quality: 79355 at least Q20
 *Consensus quality: 75365 at least Q30
 *Consensus quality: 69022 at least Q40
 Estimated insert size: agarose-PP - N/A
 **Estimated insert size: 89115 - sum-of-contigs

* 81353	83169: contig of 1817 bp in length
* 83170	83189: gap of unknown length
* 83190	84096: contig of 907 bp in length
* 84116	84116: gap of unknown length
* 84117	85749: contig of 1633 bp in length
* 85750	85769: gap of unknown length
* 85770	86770: contig of 1001 bp in length
* 86771	86790: gap of unknown length
* 86791	88207: contig of 1417 bp in length
* 88208	88227: gap of unknown length
* 88228	88976: contig of 749 bp in length
* 88977	88996: gap of unknown length
* 88997	89855: contig of 859 bp in length
FEATURES	
source	
1..89855	
/organism="Mus musculus"	
/strain="CS7BL6/J"	
/db_xref="taxon:10090"	
/chromosome="17"	
/clone="RP23-76116"	
/sex="male"	
1..79962	
misc_feature	/note="assembly_name:Contig145"
misc_feature	7983..13924
misc_feature	/note="assembly_name:Contig144"
misc_feature	13945..18590
misc_feature	/note="assembly_name:Contig143"
misc_feature	19011..22501
misc_feature	/note="assembly_name:Contig142"
misc_feature	22522..27077
misc_feature	/note="assembly_name:Contig141"
misc_feature	27098..28819
misc_feature	/note="assembly_name:Contig140"
misc_feature	28840..31224
misc_feature	/note="assembly_name:Contig139"
misc_feature	31245..335093
misc_feature	/note="assembly_name:Contig138"
misc_feature	35114..38611
misc_feature	/note="assembly_name:Contig137"
misc_feature	38632..42261
misc_feature	/note="assembly_name:Contig136"
misc_feature	42822..44130
misc_feature	/note="assembly_name:Contig135"
misc_feature	44151..47059
misc_feature	/note="assembly_name:Contig134"
misc_feature	47080..49496
misc_feature	/note="assembly_name:Contig133"
misc_feature	49517..52236
misc_feature	/note="assembly_name:Contig132"
misc_feature	52257..54270
misc_feature	/note="assembly_name:Contig131"
misc_feature	54291..56271
misc_feature	/note="assembly_name:Contig130"
misc_feature	56292..58849
misc_feature	/note="assembly_name:Contig129"
misc_feature	58870..60621
misc_feature	/note="assembly_name:Contig128"
misc_feature	60642..62255
misc_feature	/note="assembly_name:Contig127"
misc_feature	62276..63085
misc_feature	/note="assembly_name:Contig126"
misc_feature	63106..65440
misc_feature	/note="assembly_name:Contig125"
misc_feature	65461..67629
misc_feature	/note="assembly_name:Contig124"
misc_feature	67650..69178
misc_feature	/note="assembly_name:Contig123"
misc_feature	69199..70095
misc_feature	/note="assembly_name:Contig122"
misc_feature	70116..71821
misc_feature	/note="assembly_name:Contig121"
misc_feature	71842..73506
misc_feature	/note="assembly_name:Contig120"

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 21690: contig of 21690 bp in length
 * 21691 21732: gap of unknown length
 * 21733 63992: contig of 42260 bp in length
 * 63993 64034: gap of unknown length
 * 64035 73846: contig of 9812 bp in length
 * 73847 73888: gap of unknown length
 * 73889 89053: contig of 15165 bp in length
 * 89054 89095: gap of unknown length
 * 89096 97762: contig of 8667 bp in length
 * 97763 97803: gap of unknown length
 * 97804 100038: contig of 2235 bp in length
 * 100039 100079: gap of unknown length
 * 100080 109963: contig of 9884 bp in length.

FEATURES

Location/Qualifiers

1..109963
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="3"
 /clone="OSJNB0106M04"

BASE COUNT 30639 a 24707 c 24423 g 29906 t 288 others
 ORIGIN

Query Match 79.2% Score 19.8; DB 2; Length 109963;
 Best Local Similarity 91.3%; Pred. No. 60;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatgcc 24

Db 34022 GTTCAGATGGACTGCATGCC 34000

RESULT

AC011401
 LOCUS Homo sapiens chromosome 5 clone CTB-35K5, linear HTG 04-OCT-2001
 DEFINITION 11 ordered pieces.

AC011401
 AC011401.8 GI:12830130
 VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
 KEYWORDS human.
 SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 184268)

REFERENCE AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 5
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 184268)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Feb 14, 2001 this sequence version replaced gi:9954634.

COMMENT

-----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

 Project Information
 Center Project Name: 77889, H187
 Center clone name: CIT978SKB_35K5

Summary Statistics

Consensus quality: 173129 bases at least Q40
 Consensus quality: 180377 bases at least Q30
 Consensus quality: 182040 bases at least Q20
 Estimated insert size: 200000; pulse field gel estimation
 Estimated insert size: 183268; sum-of-contigs estimation
 Quality coverage: 6.62 in Q20 bases; pulse field gel estimation
 Quality coverage: 7.22 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 21632: contig of 21632 bp in length
 * 21633 21732: gap of unknown length
 * 21733 29338: contig of 7606 bp in length
 * 29339 29438: gap of unknown length
 * 29439 32379: contig of 2941 bp in length
 * 32380 32479: gap of unknown length
 * 32480 47614: contig of 15135 bp in length
 * 47615 47714: gap of unknown length
 * 47715 63279: contig of 15565 bp in length
 * 63280 63379: gap of unknown length
 * 63380 69938: contig of 6459 bp in length
 * 69939 71512: contig of 1574 bp in length
 * 71513 71612: gap of unknown length
 * 71613 77519: contig of 5907 bp in length
 * 77520 77619: gap of unknown length
 * 77620 119849: contig of 42230 bp in length
 * 119850 119949: gap of unknown length
 * 119950 146636: contig of 26687 bp in length
 * 146637 146736: gap of unknown length
 * 146737 184268: contig of 37532 bp in length.

FEATURES

source

1..184268
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTB-35K5"
 /clone_lib="CalTech human BAC library B"
 BASE COUNT 52948 a 40258 c 38925 g 51136 t 1001 others
 ORIGIN

Query Match 79.2% Score 19.8; DB 2; Length 184268;
 Best Local Similarity 91.3%; Pred. No. 61;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatgcc 24

Db 9417 GTTCAGAAATGGACTCCATGCC 9439

RESULT

AC005740/c

LOCUS Homo sapiens chromosome 5p, BAC clone 50g21 (LBNL H154), complete
 DEFINITION sequence.

AC005740

VERSION AC005740.1 GI:3687210
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 186780)
 Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
 Kadner, K., Miguel, T., Miller, C., Pittluck, S., Pollard, M.,

[illegible]

```

CDS
78, 914
/note="Vector: pOTB7"
/codon_start=1
/product="Unknown (protein for MGC:4122)"
/protein_id="AAH09561.1"
/db_xref="GI:14286320"
/translation="MDSMEPASRCLLLPDLLLLLLPAGLGPQSAEENDAVR
LPSCKEYAVELKSAPEEFGTKTEVIGTGYGLDQKASGVKYKSDLRLIEVET
ICKRLDYSILHKRTGNSRPAFGSETFTGLHNLVHKGVKVMVDIYELMNETSAEVA
DLKKQCDVLVEFEFEEVDIWRNHQDELTFLCANHLVKGKDTSCLAEGQWSKKGPT
AALGGKKKKKKSRRAKAGRRSSQKRELGLGLEDGDPSPDEDEGIQAKASPLTHSPDD
EL"
BASE COUNT      351 a    400 c    422 g    249 t
ORIGIN
Query Match      76.8%; Score 19.2; DB 9; Length 1422;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

RESULT	11	
BC004423/c		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
REMARK		
COMMENT		

```

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL plate: 9 Row: 0 Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5729791.

Location/Qualifiers
1..1720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:3530 IMAGE:2819660"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"

```

```

/lab_host="DH10B-R"
/notes="Vector: pOTB7"
388. 1224
/codon_start=1
/product="Unknown (protein for MGC:3530)"
/protein_id="AAH04423.1"
/db_xref="GI:13325208"
/translation="MDSMEPASRCLLLPLLLLLLLLPAPELGPSSQAGAEENDWVR
LPKSCVCKYVAVELKSAPEETCKTKEVIGTGYGILDKASGVKYTKSDLRLIEVTET
ICKRLDYSLHKERTGSRNRAKMSSETFEFLNLVHKGVKVMVDIPELWNETSAEVA
DLKKQDVLVEEVEEDIEDYRNHQEDLFEFLCAHVLKGDTSCLAEQNSKKGDT
AALGKKKKKKSKRAAGGRSSSKQKELGLEGDPSPDEDEGIQKASPLTHSPDD
EL"
BASE COUNT 389 a 491 c 526 g 314 t
ORIGIN
Query Match 76.8%; Score 19.2; DB 9; Length 1720;
Best Local Similarity 87.5%; Pred. No. 1.1e+07;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 gctcagacatggactccatggccc 25
||||| ||| || ||||| |||||
Db 406 GCTCAGGCAATGAATCCATGGCCCC 383

RESULT 12
LOCUS HSB20F6 18120 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone SC22CB-20F6 on chromosome 22,
complete sequence.
ACCESSION 282174
VERSION 282174.2 GI:6572311
KEYWORDS HFG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Whiteley,M. Isson
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:1695140.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
SC22CB-20F6 is from the human chromosome 22-specific cosmid library
(SC22CB) constructed at the Sanger Centre by Mark Ross and Cordelia
Langford.
VECTOR: lawlist16
IMPORTANT: This sequence is not the entire insert of clone
SC22CB-20F6 it may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap

```

```

between neighbouring submissions.
The true right end of clone SC22CB-20F6 is at 18120 in this
sequence. The true right end of clone SC22CB-79B4 is at 105 in this
sequence.
FEATURES             Location/Qualifiers
     source            1..18120
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="SC22CB-20F6"
                     /clone_l1b="SC22CB"
                     3..1189
repeat_region         /note="AluJb repeat: matches 109..281 of consensus"
                     665..792
repeat_region         /note="FLAM_C repeat: matches 1..131 of consensus"
                     820..1366
repeat_region         /note="L1Fb1 repeat: matches 5634..6155 of consensus"
                     858..862
unsure                /note="uncertain number of bases"
                     1437..1472
repeat_region         /note="L2 repeat: matches 2704..2739 of consensus"
                     1758..1950
repeat_region         /note="L2 repeat: matches 2539..2740 of consensus"
                     1942..2100
repeat_region         /note="MIR repeat: matches 48..201 of consensus"
                     2186..2208
repeat_region         /note="L2 repeat: matches 2597..2617 of consensus"
                     2209..2478
repeat_region         /note="AluJo repeat: matches 1..302 of consensus"
                     2479..2599
repeat_region         /note="L2 repeat: matches 2617..2749 of consensus"
                     3589..3718
repeat_region         /note="MIR repeat: matches 47..184 of consensus"
                     4853..5165
repeat_region         /note="Alusq repeat: matches 1..313 of consensus"
                     5313..5417
repeat_region         /note="MIR repeat: matches 45..155 of consensus"
                     5507..5629
repeat_region         /note="MIR repeat: matches 16..140 of consensus"
                     6183..6226
repeat_region         /note="L1p repeat: matches 87..130 of consensus"
                     7093..7209
repeat_region         /note="MIR repeat: matches 50..180 of consensus"
                     7772..8079
repeat_region         /note="Alusx repeat: matches 1..306 of consensus"
                     8203..8445
repeat_region         /note="L2 repeat: matches 2396..2750 of consensus"
                     8545..8841
repeat_region         /note="Alusx repeat: matches 1..293 of consensus"
                     10285..10473
repeat_region         /note="MER20 repeat: matches 1..218 of consensus"
                     10551..10606
repeat_region         /note="28 copies 2 mer ac 75 conserved"
                     10651..11166
repeat_region         /note="L1MC1 repeat: matches 5770..6315 of consensus"
                     12417..12852
repeat_region         /note="MLT1C repeat: matches 1..466 of consensus"
                     13073..13314
repeat_region         /note="Alusg repeat: matches 21..296 of consensus"
                     13496..13500
unsure                /note="uncertain number of bases"
                     13590..13667
repeat_region         /note="MIR repeat: matches 61..140 of consensus"
                     13677..13721
repeat_region         /note="MLT1C repeat: matches 1..45 of consensus"
                     13834..14027
repeat_region         /note="MLT1D repeat: matches 1..204 of consensus"
                     14069..14363
repeat_region         /note="Alusx repeat: matches 1..293 of consensus"
                     14443..15245
misc_feature          /note="match: GSS: Em:AQ747573"
                     15017..15272
repeat_region         /note="Alusx repeat: matches 35..290 of consensus"

```

```

repeat_region 15561..16022
/notes="Mfuid repeat: matches 1..503 of consensus"
repeat_region 16121..16339
/notes="L2 repeat: matches 2485..2710 of consensus"
misc_feature 16133..16577
/notes="match: GSS: Em:AQ762426"
repeat_region 17574..17885
/notes="AluX repeat: matches 1..312 of consensus"
BASE COUNT 3805 a 4734 c 4887 g 4694 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 18120;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agctcagacatgactccatggcc 24
||||| ||||| ||||| ||||| |||||
Db 11564 AGCTGAGTCGTGGACTCCATGGCC 11587

RESULT 13
HS475N16
LOCUS HS475N16 113109 bp DNA linear PRI 10-FEB-2000
DEFINITION Human DNA sequence from clone RP3-475N16 on chromosome 6p12.3-21.2.
Contains the genes for CTG4A, pre-T cell receptor alpha, a novel
protein similar to RPL7A (60S ribosomal protein L7A) and the 3' end
of gene KIAA0240. Contains ESTs, STSS, GSSs and four putative CpG
islands, complete sequence.
ACCESSION AL035587
VERSION AL035587.5 GI:6002306
KEYWORDS HTG; CpG island; CTG4A; KIAA0240; RPL7A.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 113109)
Direct Submission
Submitted (02-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 30, 1999 this sequence version replaced gi:5921383.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/
This sequence was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP3-475N16 is from the library RPI-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-475N16.
Location/Qualifiers
1..113109
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source

```



```
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgcc 24
||||| ||||| ||||| ||||| |||||
Db 26029 AGCTGATACAGGACTCCATGCC 26052

RESULT 15
AL671848 191719 bp DNA linear HTG 08-FEB-2002
LOCUS Mus musculus chromosome X clone RP23-110K5, *** SEQUENCING IN
DEFINITION
ACCESSION AL671848
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Direct Submission
Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi:18491505.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bm110K5
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 189652 bases at least Q40
Consensus quality: 190201 bases at least Q30
Consensus quality: 190528 bases at least Q20
Insert size: 191119; sum-of-contigs
Insert size: 179804; 0.5% error; agarose-fp
Quality coverage: 9.71x in Q20 bases; sum-of-contigs Quality
coverage: 10.44x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
    source
        1..191719
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="X"
            /clone="RP23-110K5"
            /clone_lib="RPCI-23"
            1..2127
            /note="assembly_fragment:03655"
            fragment_chain:1
            clone_end:T7
            vector_side:left"
            2228..40239
            /note="assembly_fragment:04272"
            fragment_chain:1"
            40340..64658
            /note="assembly_fragment:02305"
            fragment_chain:1"
            64759..81283
            /note="assembly_fragment:04002"
            fragment_chain:1"
            81384..140088
            /note="assembly_fragment:03265"
            fragment_chain:1"

misc_feature 140189..157395
            /note="assembly_fragment:04052"
            fragment_chain:1"
            157496..191719
            /note="assembly_fragment:04056"
            fragment_chain:1"
            clone_end:SP6
            vector_side:right"

BASE COUNT 54268 a 37719 c 37756 g 61366 t 610 others
ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 191719;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgcc 24
||||| ||||| ||||| ||||| |||||
Db 68165 AGCTCAGAGATGTAGTCCATGCC 68188

Search completed: May 24, 2002, 00:12:11
Job time: 11237 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 23:05:27 ; Search time 4203.75 seconds
(without alignments)
80.267 Million cell updates/sec

Title: US-09-696-686-46
Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pln:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No. Score Query Match Length DB ID Description

1 25 100.0 260 9 BB666889 BB666889 BB666889

2 25 100.0 330 9 BB856691 BB856691 BB856691

3 25 100.0 382 9 AA048794 AA048794 mj48e06.r

4 25 100.0 406 9 AA002532 AA002532 mg54b02.r

5 25 100.0 435 9 AI098985 AI098985 uc01g04.r

6 25 100.0 472 9 BB863087 BB863087 BB863087

7 25 100.0 483 9 BB857917 BB857917 BB857917

8 25 100.0 485 9 BB857913 BB857913 BB857913

9 25 100.0 495 10 BG062828 BG062828 L0958b12-

10 25 100.0 502 9 BB853356 BB853356 BB853356

11 25 100.0 608 9 BB605812 BB605812 BB605812

12 25 100.0 623 9 BB661026 BB661026 BB661026

13 25 100.0 625 9 BB633565 BB633565 BB633565

14 25 100.0 641 9 BB637089 BB637089 BB637089

15 25 100.0 643 9 BB661603 BB661603 BB661603

16 25 100.0 649 9 BB652011 BB652011 BB652011

17 25 100.0 651 10 BE289316 BE289316 601089839

18	25	100.0	725	10	BG296771	BG296771	602394543
19	25	100.0	741	10	BI081472	BI081472	602879994
20	25	100.0	776	10	BG864846	BG864846	602799277
21	25	100.0	779	10	BI688404	BI688404	603316338
22	25	100.0	790	10	BF301937	BF301937	602032966
23	25	100.0	793	10	BI104487	BI104487	602890091
24	25	100.0	1553	11	AK005532	Mus muscu	AK005532
25	24	96.0	490	9	BE226435	la20q05.y	BE226435
26	24	96.0	637	9	BB650071	BB650071	BB650071
27	24	96.0	821	10	BI653380	BI653380	603301117
28	24	96.0	1586	11	AK010702	Mus muscu	AK010702
29	23.4	93.6	249	9	AW356648	39152 MAR	AW356648
30	23.4	93.6	305	10	BE484870	171676 BA	BE484870
31	23.4	93.6	341	9	AW446522	86241 MAR	AW446522
32	23.4	93.6	379	10	BM105433	508902 MA	BM105433
33	23.4	93.6	410	10	BM287749	528547 MA	BM287749
34	23.4	93.6	411	9	AW357963	41782 MAR	AW357963
35	23.4	93.6	416	10	BF651657	274494 MA	BF651657
36	23.4	93.6	429	9	AW669666	113161 MA	AW669666
37	23.4	93.6	438	10	BG364983	284002 MA	BG364983
38	23.4	93.6	487	10	BI454088	603174992	BI454088
39	23.4	93.6	502	10	BI681908	461416 MA	BI681908
40	23.4	93.6	541	10	BE898002	187691 MA	BE898002
41	23.4	93.6	558	9	BB709265	BB709265	BB709265
42	23.4	93.6	561	9	AV612846	AV612846	AV612846
43	23.4	93.6	573	9	AV597535	AV597535	AV597535
44	21.4	85.6	222	10	BE477585	161361 BA	BE477585
45	21.4	85.6	220	9	AA296961	EST112514	AA296961

ALIGNMENTS

RESULT 1

BB666889 RIKEN full-length enriched, 260 bp mRNA linear EST 26-OCT-2001
LOCUS BB666889 RIKEN full-length enriched, 2 days pregnant adult female
DEFINITION ovary Mus musculus cDNA clone E330035012 5', mRNA sequence.
ACCESSION BB666889
VERSION BB666889.1 GI:16500522
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 260)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
TITLE Contact: Yoshihide Hayashizaki
JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
COMMENT The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE
JOURNAL
COMMENT

Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:290106
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 344.

FEATURES
source

1..382
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:479362"
/clone_lib="Soares mouse embryo NDM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TCTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT
ORIGIN

76 a 102 c 118 g 86 t
Query Match 100.0%; Score 25; DB 9; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 154 TCGATTGCCAGCAATGCCGAAGT 178

RESULT
AA002532
LOCUS

DEFINITION
mq54b02.r1 Soares mouse embryo NDM13.5 14.5 Mus musculus cDNA
clone IMAGE:427563 5', mRNA sequence.

AA002532
VERSION
AA002532.1 GI:1445971
EST.

SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 406)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT

The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:262115
Seq primer: mob.REGA+ET
High quality sequence stop: 328.

FEATURES
source

1..406
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:427563"
/clone_lib="Soares mouse embryo NDM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT
ORIGIN

81 a 107 c 128 g 90 t
Query Match 100.0%; Score 25; DB 9; Length 406;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 152 TCGATTGCCAGCAATGCCGAAGT 176

RESULT
AI098985
LOCUS

DEFINITION
AI098985 435 bp mRNA linear EST 20-AUG-1998
uc01g04.r1 Soares mammary_gland_NBMG Mus musculus cDNA clone
IMAGE:1396758 5' similar to TR:015412 O15412 CTG4A.; mRNA
sequence.

AI098985
VERSION
AI098985.1 GI:3448510
EST.
SOURCE
house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 435)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT

The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:908474

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 329.
Location/Qualifiers

FEATURES
source
1..435
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1396758"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAAGTCGGAGCGCGCGGAATGGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 85 a 110 c 134 g 106 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 435;
Best Local Similarity 100.0%; Pred. No. 0.23; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;
QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
DB 123 TGCATTGCCACCAATGCGAAGT 147

RESULT 6
BB863087 472 bp mRNA linear EST 27-NOV-2001
LOCUS
DEFINITION
CDNA Mus musculus cDNA clone G430034D12 5', mRNA sequence.
BB863087
BB863087.1 GI:17109295
EST.
house mouse.
Mus musculus

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 472)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
A., Takahashi,F., Takaku-Akaira,S., Tanaka,T., Tomaru,A., Toya,T.,
Wakahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

JOURNAL
COMMENT
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9216
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
Location/Qualifiers
1..472
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G430034D12"
/clone_lib="RIKEN full-length enriched, Leydig cells
CRL-2065 MLTC-1 cDNA"
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
92 a 127 c 156 g 97 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;
QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
DB 204 TGCATTGCCACCAATGCGAAGT 228

RESULT 7
BB857917 483 bp mRNA linear EST 26-NOV-2001
LOCUS
DEFINITION
CDNA clone G370047K11 5', mRNA sequence.
BB857917
BB857917.1 GI:17099371
EST.
house mouse.
Mus musculus

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 483)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
A., Takahashi,F., Takaku-Akaira,S., Tanaka,T., Tomaru,A., Toya,T.,
Wakahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

JOURNAL
COMMENT
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.

FEATURES
source

Location/Qualifiers

1. .483
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370047K11"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"

95 a 132 c 159 g 97 t

BASE COUNT 95 a 132 c 159 g 97 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 483;

Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaaatgcgaagt 25

|||||
 Db 220 TCGATTGCCAGCAAAATGCCGAAGT 244

RESULT 8

BB857913

LOCUS

BB857913 RIKEN full-length enriched, B16 F10Y cells Mus musculus

CDNA clone G370047K04 5', mRNA sequence.

BB857913

BB857913.1 GI:17099367

EST.

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 483)

Akimura, T., Arawaka, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii

Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,

Salto, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,

Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa

A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,

Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

2001)

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

e mouse tissues.

Location/Qualifiers

1. .485

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="G370047K04"

/clone_lib="RIKEN full-length enriched, B16 F10Y cells"

/cell_type="B16 F10Y cells"

101 a 132 c 154 g 98 t

BASE COUNT 101 a 132 c 154 g 98 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 485;

Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaaatgcgaagt 25

|||||

Db 201 TCGATTGCCAGCAAAATGCCGAAGT 225

RESULT 9

BG062828

LOCUS

BG062828 495 bp mRNA linear EST 25-JAN-2001

DEFINITION L0958D12-5 NIA Mouse Newborn Kidney cDNA Library2 (Short) Mus

musculus cDNA clone L0958D12 5', mRNA sequence.

ACCESSION BG062828

VERSION BG062828.1 GI:12533696

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 495)

Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.

Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification

method

Genome Res. 11 (9), 1553-1558 (2001)

21429098

JOURNAL

MEDLINE

COMMENT

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

Plate: L0958 row: D column: 12

Seq primer: -21M13 Reverse

High quality sequence stop: 495

POLYA-No. Location/Qualifiers

1. .495 /organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="niaEST:L0958D12-5"

/db_xref="taxon:10090"

/clone="L0958D12"

/clone_lib="NTA Mouse Newborn Kidney cDNA Library2 (Short)"

/tissue_type="Newborn Kidney"

/dev_stage="Newborn"

/lab_host="DH10B"

/note=Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a short-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). In brief, double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-PGACGATTCAGATCGGACGCGCCGCTTTT-3'] from 26 ug of total RNA, treated with 14 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-L. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 1.5 kb. The library was constructed by Yulan Piao(NIA)."

BASE COUNT 111 a 126 c 152 g 106 t

ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 495;

Best Local Similarity 100.0%; Pred. No. 0.24; Mismatches 0; Indels 0; Gaps 0;

Matches 25; Conservative 0;

QY 1 tcgcattgccagcaaatgcgaagt 25

|||||

Db 138 TCGGATTGCCAGCAAAATGCGAAGT 162

RESULT 10

BB853356

LOCUS

DEFINITION BB853356 RIKEN full-length enriched, B16 F10Y cells Mus musculus

CDNA clone G370012D12 5', mRNA sequence.

BB853356

ACCESSION

VERSION BB853356.1 GI:17094810

KEYWORDS

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 502)

REFERENCE

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayashizaki,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE

JOURNAL

COMMENT Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp/

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

e mouse tissues.

Location/Qualifiers

1. 502

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="G370012D12"

/clone_lib="RIKEN full-length enriched, B16 F10Y cells"

/cell_type="B16 F10Y cells"

BASE COUNT 103 a 136 c 161 g 102 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 502;

Best Local Similarity 100.0%; Pred. No. 0.25; Mismatches 0; Indels 0; Gaps 0;

Matches 25; Conservative 0;

QY 1 tcgcattgccagcaaatgcgaagt 25

|||||

Db 197 TCGGATTGCCAGCAAAATGCGAAGT 221

RESULT 11

BB605812

LOCUS

DEFINITION BB605812 RIKEN full-length enriched, 0 day neonate lung Mus musculus

CDNA clone E030013D02 5', mRNA sequence.

BB605812

ACCESSION

VERSION BB605812.2 GI:16450910

KEYWORDS

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 608)

REFERENCE

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Dec 5, 2000 this sequence version replaced gi:11557214.

Contact: Yoshihide Hayashizaki

dbb
146 TGCGATTGCCCGCAGCAAAATGCGAAGT 170

FEATURES
SOURCE

```

tissues:
  Location/Qualifiers
    1...623
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="DG30017H14"
      /clone_lib="RIKEN full-length enriched, 0 day neonate
        kidney"

```


contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGCGCGCCGACTCGAGTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 127 a 164 c 205 g 126 t 1 others
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.27; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

Qy 1 tcgattgcccagcaaatgcgaagt 25
|||||
Db 197 TCGATTGCCAGCAATCGGAGT 221
|||||

RESULT 13
BB633565 625 bp mRNA linear EST 26-OCT-2001
LOCUS BB633565 RIKEN full-length enriched, adult male spinal cord Mus
DEFINITION musculus cDNA clone A330062B19 5', mRNA sequence.

ACCESSION BB633565
VERSION BB633565.1 GI:16469989
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 625)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda

, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Toned, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosaka, H., Yamashita, I., Aizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

FEATURES
source

1. 625

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="A330062B19"

/clone_lib="RIKEN full-length enriched, adult male spinal
cord"

/sex="male"

/tissue_type="spinal cord"

/dev_stage="adult"

/lab_host="DH10B"

/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGGATCCGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a Lambda
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 145 a 152 c 200 g 128 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 625;
Best Local Similarity 100.0%; Pred. No. 0.27; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

Qy 1 tcgattgcccagcaaatgcgaagt 25
|||||
Db 145 TCGATTGCCAGCAATCGGAGT 169
|||||

RESULT 14
BB637089

LOCUS BB637089 RIKEN full-length enriched, adult male aorta and vein Mus
DEFINITION musculus cDNA clone A530052E13 5', mRNA sequence.

ACCESSION BB637089
VERSION BB637089.1 GI:16472874
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 641)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda

, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Toned, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosaka, H., Yamashita, I., Aizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001).
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

Location/Qualifiers
1. 625

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="A330062B19"

/clone_lib="RIKEN full-length enriched, adult male spinal
cord"

/sex="male"

/tissue_type="spinal cord"

/dev_stage="adult"

/lab_host="DH10B"

/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGGATCCGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGATCTCGAGTTAATTAATTAATCCGCCGCCGCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a Lambda
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 145 a 152 c 200 g 128 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 625;
Best Local Similarity 100.0%; Pred. No. 0.27; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

Qy 1 tcgattgcccagcaaatgcgaagt 25
|||||
Db 145 TCGATTGCCAGCAATCGGAGT 169
|||||

RESULT 14
BB637089

LOCUS BB637089 RIKEN full-length enriched, adult male aorta and vein Mus
DEFINITION musculus cDNA clone A530052E13 5', mRNA sequence.

ACCESSION BB637089
VERSION BB637089.1 GI:16472874
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 641)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda

, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 , M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 , S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 , Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa
 , K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.

FEATURES source

```

Location/Qualifiers
1. .641
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="A530052E13"
  /clone_lib="RIKEN full-length enriched, adult male aorta
  and vein"
  /sex="male"
  /tissue_type="aorta and vein"
  /dev_stage="adult"
  /lab_host="DH10B"
  /note="Site_1: Sali; Site_2: BamHI; cDNA library was
  prepared and sequenced in Mouse Genome Encyclopedia
  Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in
  RIKEN. Division of Experimental Animal Research in Riken
  contributed to prepare mouse tissues. 1st strand cDNA was
  primed with a primer [5'
  GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
  prepared by using trehalose thermo-activated reverse
  transcriptase and subsequently enriched for full-length by
  cap-trapper. cDNA went through one round of normalization
  to Rot = 20.0 and subtraction to Rot = 459.0. Second
  strand cDNA was prepared with the primer adapter of
  sequence [5' GAGAGAGAGATTCGAGTTAAATTAATACCCCCCCCCCC
  3']. cDNA was cleaved with XhoI and BamHI. Vector: a
  modified pBluescript KS(+) after bulk excision from Lambda
  FLC I."
BASE COUNT      142 a      161 c      206 g      132 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 9; Length 641;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccccagcaaatgcgaagt 25
 |||||||||||||||||||||||||
 Db 175 TGGCATTGCCCCAGCAAAATGCCGAAGT 199

RESULT 15
 BB661603

LOCUS
 DEFINITION BB661603 RIKEN full-length enriched, 0 day neonate kidney Mus
 musculus cDNA clone D630047M13 5', mRNA sequence.
 ACCESSION BB661603
 VERSION BB661603.1 GI:16495382
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 643)
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
 Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda
 , M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
 , D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
 Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 UNPUBLISHED (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 , M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 , S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 , Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa
 , K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.

FEATURES source

```

Location/Qualifiers
1. .643
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="D630047M13"
  /clone_lib="RIKEN full-length enriched, 0 day neonate
  kidney"
  /tissue_type="kidney"
  /dev_stage="0 day neonate"
  /lab_host="DH10B"
  /note="Site_1: Sali; Site_2: BamHI; cDNA library was
  prepared and sequenced in Mouse Genome Encyclopedia
  Project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in
  RIKEN. Division of Experimental Animal Research in Riken
  contributed to prepare mouse tissues. 1st strand cDNA was
  primed with a primer [5'
  GAGAGAGAGGCGCCGACCTGAGCTTTTTTTTTTTTNN 3'], cDNA was
  prepared by using trehalose thermo-activated reverse

```

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 138 a 162 c 216 g 127 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcattgccagcaaatgcgaagt 25
|||||
Db 209 TCGATTGCCCGCAGCAATGCGAAGT 233

Search completed: May 23, 2002, 23:05:31
Job time: 7487 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:18:05 ; Search time 121.39 Seconds
(without alignments)
50.588 Million cell updates/sec

Title: US-09-696-686-46
Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.4	85.6	744	4	US-09-163-285-3
2	21.4	85.6	1512	4	US-09-163-285-1
3	16.4	65.6	4195	1	US-08-200-016-1
4	16.2	64.8	1980	1	US-08-350-741-1
5	16.2	64.8	1980	2	US-08-463-875A-1
6	16	64.0	361	2	US-08-465-380-39
7	16	64.0	361	2	US-08-486-397-39
8	16	64.0	361	2	US-08-486-399-39
9	16	64.0	361	2	US-08-461-965-39
10	16	64.0	361	2	US-08-634-641-39
11	16	64.0	361	3	US-09-249-471-39
12	16	64.0	361	3	US-09-249-472-39
13	16	64.0	361	3	US-09-249-451-39
14	16	64.0	361	3	US-08-809-455-39
15	16	64.0	361	3	US-09-249-461-39
16	16	64.0	361	3	US-09-249-448-39
17	16	64.0	2941	4	US-09-425-383-1
18	16	64.0	4411529	4	US-09-103-840A-1
19	15.8	63.2	1080	3	US-09-150-133-12
20	15.8	63.2	1080	3	US-09-150-141-12
21	15.8	63.2	1080	4	US-09-374-493-12
22	15.8	63.2	1080	4	US-09-374-824-12
23	15.8	63.2	1080	4	US-09-374-492-12
24	15.6	62.4	1854	3	US-08-923-454A-13
25	15.6	62.4	4928	1	US-08-399-561-1
26	15.4	61.6	1160	2	US-08-955-138-5
27	15.4	61.6	1835	4	US-09-338-671-3

28	15.4	61.6	9542	4	US-08-968-685A-9	Sequence 9, Appli
29	15.4	61.6	36741	4	US-09-301-865-3	Sequence 3, Appli
30	15.2	60.8	2391	2	US-08-817-900-1	Sequence 1, Appli
31	15.2	60.8	2391	4	US-09-236-645-1	Sequence 1, Appli
32	15.2	60.8	3070	1	US-08-428-732-3	Sequence 3, Appli
33	15.2	60.8	3147	4	US-09-153-804-12	Sequence 12, Appli
34	15.2	60.8	3180	1	US-08-480-662-1	Sequence 1, Appli
35	15.2	60.8	3180	3	US-08-918-190-1	Sequence 1, Appli
36	15.2	60.8	3180	4	US-09-234-232-1	Sequence 1, Appli
37	15.2	60.8	3180	5	PCT-US96-09927-1	Sequence 1, Appli
38	15	60.0	993	2	US-08-245-511-57	Sequence 57, Appli
39	15	60.0	994	2	US-08-600-993A-57	Sequence 57, Appli
40	15	60.0	2464	1	US-07-863-169A-4	Sequence 4, Appli
41	15	60.0	2464	2	US-08-429-964-4	Sequence 4, Appli
42	15	60.0	2464	3	US-07-935-087-4	Sequence 4, Appli
43	15	60.0	2464	5	PCT-US93-08062-4	Sequence 4, Appli
44	15	60.0	3180	1	US-07-944-525-1	Sequence 1, Appli
45	15	60.0	4011	1	US-08-121-057-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-163-285-3
; Sequence 3, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..744
; US-09-163-285-3

Query Match 85.6%; Score 21.4; DB 4; Length 744;
Best Local Similarity 95.7%; Pred. No. 0.086;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
 |||||
 Db 97 CGCTTGGCCAGCAAAATGCGAAGT 119

RESULT 2

US-09-163-285-1
 ; Sequence 1, Application US/09163285
 ; Patent No. 6204013
 ; GENERAL INFORMATION:
 ; APPLICANT: Rhodadoust, Mehran
 ; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/163,285
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/090,398
 ; FILING DATE: June 24, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandragouras, Amy E.
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: MNI-049
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)742-4214
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1512 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 125..868
 ; US-09-163-285-1

Query Match 85.6%; Score 21.4; DB 4; Length 1512;
 Best Local Similarity 95.7%; Pred. No. 0.099;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
 |||||
 Db 221 CGCTTGGCCAGCAAAATGCGAAGT 243

RESULT 3

US-08-200-016-1
 ; Sequence 1, Application US/08200016
 ; Patent No. 5614397
 ; GENERAL INFORMATION:
 ; APPLICANT: Weissman, Irving
 ; APPLICANT: Lagasse, Eric
 ; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR MODULATING
 ; TITLE OF INVENTION: APOPTOSIS IN HEMATO LYMPHOID CELLS

NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 2200 Sand Hill Road, Suite 100
 ; CITY: Menlo Park
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/200,016
 ; FILING DATE: 22-FEB-1994
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rae-Venter, Barbara
 ; REGISTRATION NUMBER: 32,750
 ; REFERENCE/DOCKET NUMBER: 06037/003001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 854-5277
 ; TELEFAX: (415) 854-0875
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4195 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-200-016-1

Query Match 65.6%; Score 16.4; DB 1; Length 4195;
 Best Local Similarity 94.4%; Pred. No. 37;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgc 20
 |||||
 Db 1295 CTATTGCCAGCAAAATGC 1312

RESULT 4

US-08-350-741-1
 ; Sequence 1, Application US/08350741
 ; Patent No. 5804194
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGAN G.,
 ; APPLICANT: CHARLES I.G.,
 ; APPLICANT: HORMAECHE C.E.,
 ; APPLICANT: JOHNSON K.S.,
 ; APPLICANT: CHATFIELD S.N.
 ; TITLE OF INVENTION: LIVE VACCINES
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON and VANDERHYE PC
 ; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22201-4714
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/350,741
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/952,737

; FILING DATE: 09-MAY-1994
 ; APPLICATION NUMBER: US 07/952,737
 ; FILING DATE: 30-NOV-1992
 ; APPLICATION NUMBER: GB 9007194.5
 ; FILING DATE: 30-MAR-1990
 ; APPLICATION NUMBER: PCT/GB91/00484
 ; FILING DATE: 28-MAR-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WILSON, MARY J.
 ; REGISTRATION NUMBER: 32,955
 ; REFERENCE/DOCKET NUMBER: 117-158
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 816-4000
 ; TELEFAX: (703) 816-4100
 ; TELEX: 200797 NIXN UR
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1980 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 395..1822
 ; US-08-350-741-1

Query Match 64.8%; Score 16.2; DB 1; Length 1980;
 Best Local Similarity 85.7%; Pred. No. 39;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 gattgcccagcaaatgcgaag 24
 || ||||| ||||| ||||| |||||
 Db 496 GACTGCCCCAGCAGATGCCAAG 516

RESULT 5
 US-08-463-875A-1
 ; Sequence 1, Application US/08463875A
 ; Patent No. 5980907
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGAN, Gordon
 ; APPLICANT: CHARLES, Ian G.
 ; APPLICANT: HORMACHE, Carlos E.
 ; APPLICANT: JOHNSON, Kevin S.
 ; APPLICANT: CHATFIELD, Steven N.
 ; TITLE OF INVENTION: LIVE VACCINES
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON and VANDERHVE PC
 ; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22201-4714
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/463,875A
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/340,741
 ; FILING DATE: 07-DEC-1994
 ; APPLICATION NUMBER: US 07/952,737
 ; FILING DATE: 30-NOV-1992
 ; APPLICATION NUMBER: GB 9007194.5
 ; FILING DATE: 30-MAR-1990
 ; APPLICATION NUMBER: PCT/GB91/00484
 ; FILING DATE: 28-MAR-1991

; ATTORNEY/AGENT INFORMATION:
 ; NAME: WILSON, MARY J.
 ; REGISTRATION NUMBER: 32,955
 ; REFERENCE/DOCKET NUMBER: 117-158
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 816-4000
 ; TELEFAX: (703) 816-4100
 ; TELEX: 200797 NIXN UR
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1980 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 395..1822
 ; US-08-463-875A-1

Query Match 64.8%; Score 16.2; DB 2; Length 1980;
 Best Local Similarity 85.7%; Pred. No. 39;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 gattgcccagcaaatgcgaag 24
 || ||||| ||||| ||||| |||||
 Db 496 GACTGCCCCAGCAGATGCCAAG 516

RESULT 6
 US-08-465-380-39
 ; Sequence 39, Application US/08465380
 ; Patent No. 5863894
 ; GENERAL INFORMATION:
 ; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
 ; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
 ; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
 ; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
 ; APPLICANT: Peter W. Bergum
 ; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
 ; NUMBER OF SEQUENCES: 356
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; STREET: Suite 4700
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Word Perfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/465,380
 ; FILING DATE: June 5, 1995
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/326,110
 ; FILING DATE: October 18, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BIGGS, SUZANNE L.
 ; REGISTRATION NUMBER: 30,158
 ; REFERENCE/DOCKET NUMBER: 213/268
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 39:

```

; NAME/KEY: Coding Sequence
; LOCATION: 16...252
US-08-486-397-39

Query Match      64.0%; Score 16; DB 2; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY    2 gcgattgccagcaaatgcgaagt 25
       | ||||| ||||||| | | |
DB    74 GAGATTGCCAGCAAAATGAGGAAT 97

RESULT 8
US-08-486-399-39
; Sequence 39, Application US/08486399
; Patent No. 5866543
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Bergum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,399
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding sequence
; LOCATION: 16...252
US-08-486-399-39

Query Match      64.0%; Score 16; DB 2; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY    2 gcgattgccagcaaatgcgaagt 25
       | ||||| ||||||| | | |
DB    74 GAGATTGCCAGCAAAATGAGGAAT 97

RESULT 7
US-08-486-397-39
; Sequence 39, Application US/08486397
; Patent No. 5866542
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Bergum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; NUMBER OF SEQUENCES: 357
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,397
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/269
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:

```


; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
US-09-249-472-39

Query Match 64.0%; Score 16; DB 3; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 gcgattgccagcaaatgcgaagt 25
| | | | | | | | | | | | | | | | | |
Db 74 GAGATTGTCAGCAAAATGAGGAAT 97

RESULT 13
US-09-249-451-39
; Sequence 39, Application US/09249451
; Patent No. 6087487

; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroché, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,451
; FILING DATE:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
US-09-249-451-39

Query Match 64.0%; Score 16; DB 3; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 gcgattgccagcaaatgcgaagt 25
| | | | | | | | | | | | | | | | | |
Db 74 GAGATTGTCAGCAAAATGAGGAAT 97

RESULT 14
US-08-809-455-39
; Sequence 39, Application US/08809455
; Patent No. 6090916
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroché, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,455
; FILING DATE: April 17, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995

APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORGANISM: Necator americanus
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 16...252
US-08-809-455-39

Query Match 64.0%; Score 16; DB 3; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 gcgattgccagcaaatgcgaagt 25
DB 74 GAGATTGCCAGCAAAATGAGGAAT 97

RESULT 15
US-09-249-461-39
Sequence 39, Application US/09249461
Patent No. 6096877
GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo
APPLICANT: Messens, Joris Hilda Lieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Laroche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Gansemans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Berqum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
TITLE OF INVENTION: INHIBITORS AND ANTICOGULANT
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,461
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995

APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: Necator americanus
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 16...252
US-09-249-461-39

Query Match 64.0%; Score 16; DB 3; Length 361;
Best Local Similarity 79.2%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 gcgattgccagcaaatgcgaagt 25
DB 74 GAGATTGCCAGCAAAATGAGGAAT 97

Search completed: May 24, 2002, 00:18:10
Job time: 11556 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:23:09 ; Search time 525.6 seconds
(without alignments)
81.664 Million cell updates/sec

Title: US-09-696-686-46
Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Capext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	22	AA05281
2	25	100.0	25	22	AA05288
3	25	100.0	49	22	AA05283
4	25	100.0	1839	22	AA05278
5	25	100.0	1848	22	AA05286
6	24	96.0	370	22	AA05285
7	21.4	85.6	238	21	AA04529
8	21.4	85.6	371	20	AA088062
9	21.4	85.6	443	20	AA088063

10	21.4	85.6	510	22	AA093625
11	21.4	85.6	744	21	AA049678
12	21.4	85.6	746	22	AA041104
13	21.4	85.6	747	22	AA090704
14	21.4	85.6	1306	20	AA024902
15	21.4	85.6	1312	22	AA05462
16	21.4	85.6	1315	22	AA05464
17	21.4	85.6	1419	22	AA093914
18	21.4	85.6	1457	20	AA078923
19	21.4	85.6	1462	21	AA016621
20	21.4	85.6	1470	21	AA096349
21	21.4	85.6	1470	21	AA058116
22	21.4	85.6	1470	22	AA046164
23	21.4	85.6	1512	21	AA049677
24	21.4	85.6	1623	20	AA088064
25	20.2	80.8	1362	22	AA05279
26	20.2	80.8	1675	22	AA021311
27	20.2	80.8	10140	22	AA049201
28	19.2	76.8	1657	22	AA040377
29	18.8	75.2	568	22	AA013480
30	18.8	75.2	2096	22	AA016730
31	18.8	75.2	2096	22	AA010871
32	18.2	72.8	107	21	AA023001
33	18.2	72.8	541	22	AA092573
34	18.2	72.8	1285	22	ABL17251
35	18.2	72.8	2678	22	AA04725
36	18.2	72.8	4321	23	ABL17250
37	18.2	72.8	4487	21	AA076070
38	17.2	68.8	3154	23	ABL18228
39	17.2	68.8	18748	23	ABL05146
40	17	68.0	152	23	AA050725
41	17	68.0	377	18	AA078311
42	17	68.0	424	18	AA078020
43	17	68.0	1071	23	AA051969
44	17	68.0	1080	23	AA054722
45	17	68.0	3899	23	ABL1747

ALIGNMENTS

RESULT 1
AA05281
ID AA05281 standard; DNA; 25 BP.
XX
AC AA05281;
XX
DT 07-SEP-2001 (first entry)
XX
DE PCR primer #2 used to amplify T243 homologous sequences.
XX
KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; PCR primer; ss.
XX
OS Mus sp.
XX
PN WO200130798-A1.
XX
PD 03-MAY-2001.
XX
PF 26-OCT-2000; 2000WO-US29382.
XX
PR 26-OCT-1999; 99US-0161488.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Klein R, Matthews W, Moore M, Allen KD;
XX
DR WPI; 2001-300473/31.
XX
PT Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs).

Umbilical vein end
Human myocardium s
Primer specific fo
Human secretory pr
Human secreted pro
Human secreted pro
Human secreted pro
Human cDNA encodin
Human zcalc-1 DNA
Human secreted pro
cDNA encoding a no
Human PRO4354 nucl
Human DNA encoding
Human myocardium s
Human zcalc-1 DNA
Human trinucleotid
Human cDNA sequenc
Human CHOT promote
Human secreted pro
Human cDNA clone (c
Human cDNA sequenc
Human genomic DNA
Human secreted pro
Human cDNA 3'-end
Drosophila melanog
Human full-length
Drosophila melanog
Human ORFX ORF1625
Drosophila melanog
Drosophila melanog
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur

PT contains heterozygous disruption in a gene encoding TRP -

XX Claim 14; Fig 13; 106pp; English.

PS The present sequence for PCR primer #2 is used to amplify the T243 gene from an R1 mouse embryonic stem cell genomic library. The T243 gene encodes for a trinucleotide repeat protein (TRP). The invention describes methods of producing embryonic stem (ES) cells comprising a heterozygous disruption in a target DNA sequence (preferably gene T243) encoding a TRP and of producing a knockout mouse comprising a homozygous disruption in a gene where the disruption inhibits the production of the wild type TRP. The invention also relates to identifying agents capable of affecting a phenotypic change. The transgenic animals and the cells are useful for identifying compounds capable of ameliorating disease symptoms, and as test substrates for the identification of drugs, pharmaceuticals, therapies and interventions which may be effective in treating trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's disease. The animal models for trinucleotide repeat disorders are ideal model systems to study the progression of disease in vivo, the molecular basis of these diseases and show the features observed in human disease. Using the mice, it is possible to model both the pathogenic mechanism and the trinucleotide repeat instability in the mouse.

XX SQ Sequence 25 BP; 7 A; 6 C; 7 G; 5 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0048;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
DB 1 tgcgattgccagcaaatgcgaagt 25

RESULT 2
AAS05288/c
ID AAS05288 standard; DNA; 25 BP.

XX AAS05288;
XX 07-SEP-2001 (first entry)

XX Mouse ES cell T243 gene PCR primer #432.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX transgenic animal; knockout mouse; triplet repeat expansion;
XX mouse; PCR primer; ss.

XX Mus sp.
XX WO200130798-A1.
XX 03-MAY-2001.
XX 26-OCT-2000; 2000WO-US29382.
XX 26-OCT-1999; 99US-0161488.
XX 26-OCT-1999; 99US-0161488.
XX (DELT-) DELTAGEN INC.
XX Klein R, Matthews W, Moore M, Allen KD;
XX WPI; 2001-300473/31.
XX Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs), contains heterozygous disruption in a gene encoding TRP -

XX Example 12; Fig 13; 106pp; English.

PS The present sequence for PCR primer #489 is used to further amplify a T243 gene PCR product from pool A7 (obtained by PCR of the T243 gene

XX SQ Sequence 25 BP; 5 A; 7 C; 6 G; 7 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0048;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
DB 25 TGCgATTGCCAGCAAAATGCGAAGT 1

RESULT 3
AAS05283
ID AAS05283 standard; DNA; 49 BP.

XX AAS05283;
XX 07-SEP-2001 (first entry)

XX PCR primer #489 used to amplify T243 PCR product from pool A7.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX transgenic animal; knockout mouse; triplet repeat expansion;
XX fragile X syndrome; Huntington's disease; mouse; PCR primer; ss.

XX Mus sp.
XX WO200130798-A1.
XX 03-MAY-2001.
XX 26-OCT-2000; 2000WO-US29382.
XX 26-OCT-1999; 99US-0161488.
XX (DELT-) DELTAGEN INC.
XX Klein R, Matthews W, Moore M, Allen KD;
XX WPI; 2001-300473/31.
XX Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs), contains heterozygous disruption in a gene encoding TRP -

XX Example 12; Fig 13; 106pp; English.

PS The present sequence for PCR primer #489 is used to further amplify a T243 gene PCR product from pool A7 (obtained by PCR of the T243 gene

CC from an R1 mouse embryonic stem cell genomic library). Gene T243 encodes
 CC for a trinucleotide repeat protein (TRP). The invention describes methods
 CC of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably gene T243) encoding a
 CC TRP and of producing a knockout mouse comprising a homozygous disruption
 CC in a gene encoding TRP, where the disruption inhibits the production of
 CC the wild type TRP. The invention also relates to identifying agents
 CC capable of affecting a phenotype of a knockout mouse. The transgenic
 CC animals and the cells are useful for identifying compounds capable of
 CC ameliorating disease symptoms, and as test substrates for the
 CC identification of drugs, pharmaceuticals, therapies and interventions
 CC which may be effective in treating trinucleotide repeat disorders e.g.
 CC fragile X syndrome and Huntington's disease. The animal models for
 CC trinucleotide repeat disorders are ideal model systems to study the
 CC progression of disease in vivo, the molecular basis of these diseases
 CC and show the features observed in human disease. Using the mice, it is
 CC possible to model both the pathogenic mechanism and the trinucleotide
 CC repeat instability in the mouse.
 XX
 SQ Sequence 49 BP; 9 A; 13 C; 14 G; 13 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 49;
 Best Local Similarity 100.0%; Pred. No. 0.0054;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 25 tgcgattgccagcaaatgcgaagt 49

RESULT 4
 AAS05278
 ID AAS05278 standard; cDNA; 1839 BP.
 XX
 AC AAS05278;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Murine trinucleotide repeat protein (TRP) cDNA sequence.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; murine; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 55..885
 FT /*tag= a
 FT /product= "TRP"
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELTA-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPI; 2001-300473/31.
 DR P-PSDB; AAU02498.
 XX
 PT Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -
 XX Claim 8; Fig 11; 106pp; English.
 XX

CC The present sequence encodes for murine trinucleotide repeat protein
 CC (TRP). The invention describes methods of producing embryonic stem (ES)
 CC cells comprising a heterozygous disruption in a target DNA sequence
 CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
 CC comprising a homozygous disruption in a gene encoding TRP, where the
 CC disruption inhibits the production of the wild type TRP. The invention
 CC also relates to identifying agents capable of affecting a phenotype of
 CC a knockout mouse. Also described are methods of determining whether
 CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
 CC phenotypic change. The transgenic animals and the cells are useful for
 CC identifying compounds capable of ameliorating disease symptoms, and as
 CC test substrates for the identification of drugs, pharmaceuticals,
 CC therapies and interventions which may be effective in treating
 CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
 CC disease. The animal models for trinucleotide repeat disorders are ideal
 CC model systems to study the progression of disease in vivo, the molecular
 CC basis of these diseases and show the features observed in human disease.
 CC Using the mice, it is possible to model both the pathogenic mechanism and
 CC the trinucleotide repeat instability in the mouse.
 XX
 SQ Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 1839;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgcgattgccagcaaatgcgaagt 25
 |||||
 Db 182 tgcgattgccagcaaatgcgaagt 206

RESULT 5
 AAS05286
 ID AAS05286 standard; DNA; 1848 BP.
 XX
 AC AAS05286;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE DNA sequence of an expanded T243 gene.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; mutant; ds.
 XX
 OS Mus sp.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 55..894
 FT /*tag= a
 FT /product= "Expanded TRP"
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELTA-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPI; 2001-300473/31.
 DR P-PSDB; AAU02500.
 XX
 PT Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -
 XX

PS Disclosure; Fig 15; 106pp; English.

XX The present sequence represents the DNA sequence of an expanded T243

CC gene. The T243 gene encodes for a trinucleotide repeat protein (TRP).

CC The invention describes methods of producing embryonic stem (ES)

CC cells comprising a heterozygous disruption in a target DNA sequence

CC (preferably gene T243) encoding a TRP and of producing a knockout mouse

CC comprising a homozygous disruption in a gene encoding TRP, where the

CC disruption inhibits the production of the wild type TRP. The invention

CC also relates to identifying agents capable of affecting a phenotype of

CC a knockout mouse. Also described are methods of determining whether

CC expansion of the trinucleotide repeat in a gene encoding TRP produces a

CC phenotypic change. The transgenic animals and the cells are useful for

CC identifying compounds capable of ameliorating disease symptoms, and as

CC test substrates for the identification of drugs, pharmaceuticals,

CC therapies and interventions which may be effective in treating

CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's

CC disease. The animal models for trinucleotide repeat disorders are ideal

CC model systems to study the progression of disease in vivo, the molecular

CC basis of these diseases and show the features observed in human disease.

CC Using the mice, it is possible to model both the pathogenic mechanism and

CC the trinucleotide repeat instability in the mouse.

XX

SQ Sequence 1848 BP; 415 A; 519 C; 543 G; 371 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 1848;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaagt 25

|||||

DB 191 tgcgattgccagcaaatgcgaagt 215

RESULT 6

AA505285

ID AAS05285 standard; DNA; 370 BP.

XX

AC AAS05285;

XX

DT 07-SEP-2001 (first entry)

DE

DE T243 gene homologous sequence #2 generated by PCR.

XX

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;

XX transgenic animal; knockout mouse; triplet repeat expansion;

XX fragile X syndrome; Huntington's disease; mouse; ds.

XX

OS Mus sp.

XX

PN WO200130798-A1.

XX

PD 03-MAY-2001.

XX

PF 26-OCT-2000; 2000WO-US29382.

XX

PR 26-OCT-1999; 99US-0161488.

XX

PA (DELTA-) DELTAGEN INC.

XX

PI Klein R, Matthews W, Moore M, Allen KD;

XX

DR WPI; 2001-300473/31.

XX

XX Novel transgenic animals useful as animal model for characterization of

PT function of a gene encoding trinucleotide repeat proteins (TRPs),

PT contains heterozygous disruption in a gene encoding TRP -

XX

PS Claim 11; Fig 14; 106pp; English.

XX

XX The present sequence for T243 homologous sequence #2 is generated by PCR

CC and is homologous to the T243 gene which encodes for a trinucleotide

CC repeat protein (TRP). The invention describes methods of producing

CC embryonic stem (ES) cells comprising a heterozygous disruption in a

CC target DNA sequence (preferably gene T243) encoding a TRP and of

CC producing a knockout mouse comprising a homozygous disruption in a gene

CC encoding TRP, where the disruption inhibits the production of the wild

CC type TRP. The invention also relates to identifying agents capable of

CC affecting a phenotype of a knockout mouse. Also described are methods of

CC determining whether expansion of the trinucleotide repeat in a gene

CC encoding TRP produces a phenotypic change. The transgenic animals and the

CC cells are useful for identifying compounds capable of ameliorating

CC disease symptoms, and as test substrates for the identification of drugs,

CC pharmaceuticals, therapies and interventions which may be effective in

CC treating trinucleotide repeat disorders e.g. fragile X syndrome and

CC Huntington's disease. The animal models for trinucleotide repeat

CC disorders are ideal model systems to study the progression of disease in

CC vivo, the molecular basis of these diseases and show the features

CC observed in human disease. Using the mice, it is possible to model both

CC the pathogenic mechanism and the trinucleotide repeat instability in the

CC mouse.

XX

SQ Sequence 370 BP; 73 A; 86 C; 149 G; 62 T; 0 other;

Query Match 96.0%; Score 24; DB 22; Length 370;

Best Local Similarity 100.0%; Pred. No. 0.025;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaag 24

|||||

DB 1 tgcgattgccagcaaatgcgaag 24

RESULT 7

AAA45529

ID AAA45529 standard; cDNA; 238 BP.

XX

AC AAA45529;

XX

DT 21-AUG-2000 (first entry)

XX

DE Human secreted expressed sequence tag SEQ ID NO:2104.

XX

XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;

XX expressed sequence tag; EST; probe; chemotactic; proliferative;

XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;

XX antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;

XX antiulcer; osteopathic; neuroprotective; neurotropic; antipsoriatic;

XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;

XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;

XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

XX central nervous system disorder; Alzheimer's disease; stroke;

XX parkinson's disease; Huntington's disease; coagulation disorder;

XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

XX tumour; infection; depression; psoriasis; ss.

XX

OS Homo sapiens.

XX

PN WO200021991-A1.

XX

PD 20-APR-2000.

XX

PF 15-OCT-1999; 99WO-US24208.

XX

PR 15-OCT-1998; 98US-0104436.

XX

PA (GEM) GENETICS INST INC.

XX

PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Bowman MR;

XX

DR WPI; 2000-317938/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
XX
PS Claim 1; Page 729; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (SESTs). Isolated from human, mouse, chicken and rat
CC tissue sources, the SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytotaxtic; antibacterial; antifungal; antiviral; antidiabetic;
CC antistatic; vulnary; antitumor; osteoporotic; neuroprotective;
CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidiabetic. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 238 BP; 51 A; 51 C; 80 G; 56 T; 0 other;

Query Match 85.6%; Score 21.4; DB 21; Length 238;
Best Local Similarity 95.7%; Pred. No. 0.46;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
|| |||||
DB 155 cgcttgcagcaaatgcgaagt 177

RESULT 8
AA88062
ID AA88062 standard; DNA; 371 BP.

AC AA88062;
XX
XX 09-SEP-1999 (first entry)
XX
XX Human zcalc-1 DNA allele 1.

XX Calcitonin-like protein 1; zcalc-1; human; osteopathic; antidiabetic;
KW hypotensive; vasodilator; neuromodulator; peripheral organ; treatment;
KW Paget's disease; hypercalcaemia; osteoporosis; Raynaud's disease;
KW type I diabetes; hypertension; ss.

XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 1..371
FT /tag= a
FT /product= "zcalc-1"
FT /note= "calcitonin-like protein 1"

XX WO9931131-A2.
XX
XX 24-JUN-1999.
XX
XX 18-DEC-1998; 98WO-US26940.
XX
XX 18-DEC-1997; 97US-0993935.

XX (ZYMO) ZYMOGENETICS INC.
PA Moore EE, Raymond FC, Sheppard PO;
PI WPI; 1999-418753/35.
XX P-PSDB: AAY25418.
XX
XX New mammalian nucleic acid encoding a calcitonin-like polypeptide-1
PS Disclosure; Page 47-48; 55pp; English.
XX
XX This invention describes a novel human polynucleotide that encodes a
CC calcitonin-like polypeptide which has osteopathic, antidiabetic and
CC hypotensive activity and also acts as a vasodilator. Antibodies derived
CC from the product of the invention can be used to isolate and purify
CC the polypeptide sequences. The polypeptides can be used as
CC neuromodulators in a variety of peripheral organs. They can be used in
CC the treatment of Paget's disease, hypercalcaemia, and osteoporosis. They
CC can also be used to treat Raynaud's disease, and may be able to inhibit
CC the progression of type I diabetes. The polypeptide sequences may be
CC administered as vasodilators to treat hypertension.
XX
SQ Sequence 371 BP; 79 A; 81 C; 125 G; 86 T; 0 other;

Query Match 85.6%; Score 21.4; DB 20; Length 371;
Best Local Similarity 95.7%; Pred. No. 0.5;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
|| |||||
DB 222 cgcttgcagcaaatgcgaagt 244

RESULT 9
AA88063
ID AA88063 standard; DNA; 443 BP.

AC AA88063;
XX
XX 09-SEP-1999 (first entry)
XX
XX Human zcalc-1 DNA allele 2.

XX Calcitonin-like protein 1; zcalc-1; human; osteopathic; antidiabetic;
KW hypotensive; vasodilator; neuromodulator; peripheral organ; treatment;
KW Paget's disease; hypercalcaemia; osteoporosis; Raynaud's disease;
KW type I diabetes; hypertension; ss.

XX Homo sapiens.
XX
XX WO9931131-A2.
XX
XX 24-JUN-1999.
XX
XX 18-DEC-1998; 98WO-US26940.
XX
XX 18-DEC-1997; 97US-0993935.

XX (ZYMO) ZYMOGENETICS INC.
XX
XX Moore EE, Raymond FC, Sheppard PO;
XX WPI; 1999-418753/35.
XX
XX New mammalian nucleic acid encoding a calcitonin-like polypeptide-1
PS Disclosure; Page 48; 55pp; English.

XX This invention describes a novel human polynucleotide that encodes a
CC calcitonin-like polypeptide which has osteopathic, antidiabetic and
CC hypotensive activity and also acts as a vasodilator. Antibodies derived

CC from the product of the invention can be used to isolate and purify
CC the polypeptide sequences. The polypeptides can be used as
CC neuromodulators in a variety of peripheral organs. They can be used in
CC the treatment of Paget's disease, hypercalcaemia, and osteoporosis. They
CC can also be used to treat Raynaud's disease, and may be able to inhibit
CC the progression of type I diabetes. The polypeptide sequences may be
CC administered as vasodilators to treat hypertension.
XX
XX

XX Sequence 443 BP; 113 A; 94 C; 142 G; 94 T; 0 other;

Query Match 85.6%; Score 21.4; DB 20; Length 443;
Best Local Similarity 95.7%; Pred. No. 0.52;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaagt 25
|| |||||
DB 182 cgcttgcccagcaaatgcgaagt 204
|| |||||

RESULT 10
AAAF93625
ID AAF93625 standard; cDNA; 510 BP.
XX
AC AAF93625;
XX
XX 21-MAY-2001 (first entry)
DT

DE Umbilical vein endothelial cell cDNA encoding SRT protein SEQ ID 446.
XX
DE Human; SRT; gene therapy; gene mapping; tissue typing; ss.
KW Homo sapiens.
KW
XX
OS WO200107611-A2.
PN
XX
XX 01-FEB-2001.
PD
XX
XX 21-JUL-2000; 2000WO-US20006.
PF
XX
XX 26-JUL-1999; 99US-0145701.
PR
XX
XX (GETH) GENENTECH INC.
PA
XX
XX Baker KP, Goddard A, Wood WI;
PI
XX
XX WPI; 2001-112729/12.
DR

XX New isolated nucleic acid molecule encoding a SRT polypeptide is useful
XX for production of recombinant SRT polypeptides, gene mapping,
XX diagnosing genetic disorders and for gene therapy -
XX
XX Claim 2; Fig 446; 663pp; English.

XX Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
XX human SRT proteins. The cDNA sequences are isolated from various
XX different human tissue cDNA libraries. The invention relates to a method
XX for detecting cDNA encoding an SRT protein, a vector containing cDNA
XX encoding SRT, a host cell transformed with the vector, an isolated SRT
XX polypeptide, and an antibody which binds to SRT. The polynucleotide
XX sequence can be used in gene therapy and is useful in the recombinant
XX production of SRT polypeptides, as a hybridisation probe to screen
XX libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
XX map the gene encoding the SRT polypeptides and analysing genetic
XX disorders, tissue typing and disease tissue detection. The SRT
XX polynucleotide sequences can be used in polymerase chain reaction,
XX screening for new therapeutic molecules and generation of antisense RNA
XX and DNA.

XX Sequence 510 BP; 122 A; 114 C; 168 G; 104 T; 2 other;
XX
XX Query Match 85.6%; Score 21.4; DB 22; Length 510;

Best Local Similarity 95.7%; Pred. No. 0.53;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 cgattgccagcaaatgcgaagt 25
|| |||||
DB 206 cgcttgcccagcaaatgcgaagt 228
|| |||||

RESULT 11
AAZ49678
ID AAZ49678 standard; cDNA; 744 BP.
XX
AC AAZ49678;
XX
XX 07-APR-2000 (first entry)
DT

XX Human myocardium secreted protein-5 coding region.
DE
XX
XX Human; myocardium secreted protein-5; MSP-5; hypotensive; cardiant;
KW cardiac cellular process; cardiovascular disorder; cardiomyopathy;
KW congestive heart failure; hypertension; atherosclerosis;
KW coronary artery disease; valvular disease; arrhythmia; ss.
XX
XX Homo sapiens.
OS

XX
XX Key Location/Qualifiers
FH 1..744
FT mat_peptide /*tag= a /product= "MSP-5"
FT /note= "Myocardium secreted protein-5"
FT
FT sig_peptide 1..66
FT /*tag= b
FT mat_peptide 67..744
FT /*tag= c
FT /label= Mature_MSP-5

XX W09967385-A1.
XX
XX 29-DEC-1999.
XX
XX 22-JUN-1999; 99WO-US13937.
XX
XX 23-JUN-1998; 98US-0090398.
XX 29-SEP-1998; 98US-0163285.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Khodadoust MM;
PI
XX
XX WPI; 2000-136983/12.
XX P-PSDB; AAY44629.
XX

XX Novel myocardium secreted protein-5 polynucleotides, used to modulate a
XX variety of cellular processes -
XX
XX Claim 1; Pages 92-93; 99pp; English.

XX The present sequence is the coding region of myocardium secreted protein
XX -5 (MSP-5) cDNA which was isolated from a cDNA library prepared from a
XX cardiac tissue sample obtained from a biopsy of a 42 year old woman
XX suffering from congestive heart failure. The MSP-5 protein has
XX hypertensive and cardiant activities. It is highly expressed in heart,
XX brain, placenta, fetal lung, liver, kidney, testis, small intestine and
XX pituitary gland. The present sequence is used to modulate a variety of
XX cellular processes, especially cardiac cellular processes. MSP-5 is used
XX to modulate the activity of one or more proteins involved in a
XX cardiovascular disorder, e.g. congestive heart failure or cardiomyopathy.
XX Conditions and diseases which can be treated include hypertension,
XX atherosclerosis, coronary artery spasm, coronary artery disease, valvular
XX disease, arrhythmias, and cardiopathies (e.g. hypertrophic, dilative, or
XX restrictive cardiomyopathies), and disorders related to under or over
XX expression of MSP-5.

This invention relates to nucleic acid sequences AAF93744 - AAF93916 which encode human secretory or membrane proteins represented by AAB88317 - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the invention. The invention also includes methods for the production of antibodies directed against the proteins, and cDNA sequences, which can be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA). Examples of diseases which may be treated include rheumatoid arthritis and diabetes.

RESULT 14
AAZ24902
ID AA224902 standard; DNA; 1306 BP.
XX
XX
AC AAZ24902;
XX
DT 02-DEC-1999 (first entry)
XX
DE Human secreted protein gene 92 clone HATAC53.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS
XX Homo sapiens.
XX
PN WO9947540-A1.
XX
XX 23-SEP-1999.
PD
XX
PF 18-MAR-1999; 99WO-US05804.
XX
XX 19-MAR-1998; 98US-0078563.
PR
XX 19-MAR-1998; 98US-0078566.
PR
XX 19-MAR-1998; 98US-0078573.
PR
XX 19-MAR-1998; 98US-0078574.
PR
XX 19-MAR-1998; 98US-0078576.
PR
XX 19-MAR-1998; 98US-0078577.
PR
XX 19-MAR-1998; 98US-0078578.
PR
XX 19-MAR-1998; 98US-0078579.
PR
XX 19-MAR-1998; 98US-0078581.
PR
XX 01-APR-1998; 98US-0080312.
PR
XX 01-APR-1998; 98US-0080313.
PR
XX 01-APR-1998; 98US-0080314.
PR
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;
PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
PI Olsen HS, Shi Y, Moore PA;
XX
XX WPI; 1999-562050/47.
DR
XX P-PSDB; AAY41399.
XX
XX New isolated human genes, useful for diagnosis and treatment of e.g.
XX cancers, neurological disorders, immune diseases, inflammation or blood
XX disorders -
XX
XX Claim 1; Page 353; 484pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a
XX secreted human protein. The gene number, and the clone it is derived
XX from, are detailed in the descriptor line. The gene can be used to
XX generate fusion proteins by linking to the gene to a human immunoglobulin
XX Fc portion (e.g. AAZ24802) for increasing the stability of the fused
XX protein as compared to the human protein only.
XX The invention relates to 95 novel genes and their fragments (nucleic
XX acid sequences: AAZ24811-224907; amino acid sequences AAY41308-Y41404)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in
XX the new polynucleotides. Specific uses are described for each of the 95
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AAZ24811 for described uses).
XX
XX Sequence 1306 BP; 334 A; 291 C; 377 G; 300 T; 4 other;

Query Match 85.6%; Score 21.4; DB 20; Length 1306;
Best Local Similarity 95.7%; Pred. No. 0.63;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 cgattgcccagcaaatgcgaagt 25
|| ||||| ||||| ||||| |||||
Db 195 cgcttgcccagcaaatgcgaagt 217
RESULT 15
AAD05462
ID AAD05462 standard; cDNA; 1312 BP.
XX
XX AAD05462;
XX
DT 17-JUL-2001 (first entry)
XX
DE Human secreted protein-encoding gene 23 cDNA clone HNTRS57, SEQ ID NO:84.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; kidney disorder;
KW cardiovascular disorder; angioinfectious disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification; ss.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 69..674
XX FT /*tag= a
XX FT /product= "Human secreted protein"
XX FT /note= "CDS does not include start codon"
XX FT /partial
XX FT 69..71
XX FT sig_peptide /*tag= b
XX FT 72..671
XX FT mat_peptide /*tag= c
XX FT /product= "Mature human secreted protein"
XX
XX W0200134623-A1.
XX
XX PD 17-MAY-2001.
XX
XX 01-NOV-2000; 2000WO-US30037.
XX
XX 05-NOV-1999; 99US-0163577.
XX
XX 30-JUN-2000; 2000US-0215137.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Moore PA;
XX
XX WPI; 2001-316490/33.
XX
XX P-PSDB; AAE01619.
XX
XX Nucleic acids encoding 29 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX Claim 1; Page 469; 535pp; English.
XX
XX AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
XX protein genes, and AAE01546-AAE01630 represent the proteins they encode.
XX AAE01631-AAE01660 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing,

CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 29 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiotensin disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein-encoding cDNA of the invention.

xx
 SO Sequence 1312 BP: 334 A; 288 C; 382 G; 304 T; 4 other;

Query Match 85.6%; Score 21.4; DB 22; Length 1312;
 Best Local Similarity 95.7%; Pred. No. 0.63;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cgattgccagcaaatgcgaagt 25
 |||
 Db 201 cgcttgccagcaaatgcgaagt 223

Search completed: May 24, 2002, 00:23:11
 Job time: 9008 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:12:11 ; Search time 3953.4 Seconds
(without alignments)
132.333 Million cell updates/sec

Title: US-09-696-686-46

Perfect score: 25
Sequence: 1 tgcgattgccagcaaatgcgaagt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description

1	25	100.0	1882	10	BC013549	BC013549 Mus muscu
2	25	100.0	1909	10	AF361644	AF361644 Mus muscu
3	24	96.0	89855	2	AC074220	AC074220 Mus muscu
4	21.4	85.6	510	6	AX079702	AX079702 Sequence
5	21.4	85.6	744	6	AR142811	AR142811 Sequence
6	21.4	85.6	746	6	AX136616	AX136616 Sequence
7	21.4	85.6	747	6	BD006701	BD006701 Novel pol
8	21.4	85.6	1413	6	BC019903	BC019903 Homo sapi
9	21.4	85.6	1419	6	AX136423	AX136423 Sequence
10	21.4	85.6	1512	6	AR142810	AR142810 Sequence
11	20.4	81.6	128185	2	AC087144	AC087144 Mus muscu
12	20.4	81.6	132764	2	AC073842	AC073842 Homo sapi
13	20.4	81.6	167854	2	AC092483	AC092483 Homo sapi
14	20.4	81.6	209958	2	AC087135	AC087135 Mus muscu
15	20.2	80.8	962	9	HS080744	U80744 Homo sapien
16	20.2	80.8	1422	9	BC008898	BC008898 Homo sapi
17	20.2	80.8	1422	9	BC008961	BC008961 Homo sapi
18	20.2	80.8	1720	9	BC008963	BC008963 Homo sapi
19	20.2	80.8	190043	9	BC004423	AL035587 Human DNA
20	19.2	76.8	113109	9	HS475N16	AL035587 Human DNA
21	19.2	76.8	160609	2	AC096211	AC096211 Rattus no
22	18.8	75.2	1363	9	HUMBINDA	L10403 Homo sapien
23	18.8	75.2	2008	9	AK023954	AK023954 Homo sapi
24	18.8	75.2	2096	6	AX192223	AX192223 Sequence
25	18.8	75.2	3492	10	AF175967	AF175967 Mus muscu
26	18.8	75.2	4349	10	AF175968	AF175968 Mus muscu
27	18.8	75.2	154508	2	AC083974	AC083974 Homo sapi
28	18.8	75.2	160741	9	AC097709	AC097709 Homo sapi
29	18.8	75.2	168741	2	AC013573	AC013573 Homo sapi
30	18.8	75.2	16891	2	AC106523	AC106523 Rattus no
31	18.8	75.2	178608	2	AC027531	AC027531 Homo sapi
32	18.8	75.2	193625	2	AF001194	AF001194 Homo sapi
33	18.8	75.2	280810	2	AC046188	AC046188 Homo sapi
34	18.6	74.4	2029	10	AF184981	AF184981 Mus muscu
35	18.6	74.4	185893	9	AC005191	AC005191 Homo sapi
36	18.2	72.8	3191	9	BC011454	BC011454 Homo sapi
37	18.2	72.8	3542	9	AF175966	AF175966 Homo sapi
38	18.2	72.8	4412	9	AB023206	AB023206 Homo sapi
39	18.2	72.8	8509	2	AC012985	AC012985 Drosophila
40	18.2	72.8	28928	3	CEZC410	268270 Caenorhabdit
41	18.2	72.8	45978	3	AF016449	AF016449 Caenorhab
42	18.2	72.8	74484	2	AC106387	AC106387 Rattus no
43	18.2	72.8	92969	2	AC010844	AC010844 Drosophila
44	18.2	72.8	113635	2	AC108252	AC108252 Rattus no
45	18.2	72.8	117136	2	AC015940	AC015940 Homo sapi

ALIGNMENTS

RESULT 1

BC013549

LOCUS

DEFINITION

BC013549

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC013549 Mus musculus, RIKEN cDNA 1600025D17 gene, clone MGC:19340
IMAGE:4222133, mRNA, complete cds.

BC013549 1882 bp mRNA linear ROD 07-SEP-2001

BC013549.1 GI:15488835

MGC house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1882)

Strausberg R.

Direct Submission

Submitted (04-SEP-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: cgapps@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

Fri May 24 10:49:39 2002

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 25 Row: j Column: 9.

Location/Qualifiers
1. .1882
/organism="Mus musculus"
/db_xref="LocusID:72029"
/db_xref="taxon:10090"
/clone="MGC:19340 IMAGE:4222133"
/tissue="kidney, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
91. .921
/codon_start=1
/product="RIKEN cDNA 1600025D17 gene"
/protein_id="AAH13549.1"
/db_xref="GI:15488836"
/translation="MESSELAAPRCCLFLLPRLPAPKLGSPAGAEETDWR
LPSCVEVKYAVELKSAFETGKTKEVIDTGYGLDGKSGVYKTSDLRIEVTET
ICKRLDYSLHKERTGNSRFAGKMSSETFETLHNLVHKGKVMVDIPYELNMTSAEVA
DLKKQCDLVVEEFEEVDWYRNHQQEEDLTFELCANHLVKGKDTSCLAERWSGKGGDI
ASLGKSKKKRSGVKGSSSGSKORKELGGLGEDANAEEREGVQKASPLPHSPDDEL
"

BASE COUNT 432 a 539 c 549 g 362 t
ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1882;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 218 TCGATTGCCAGCAAAATGCGAAGT 242

RESULT 2
AF361644 1909 bp mRNA linear ROD 12-MAY-2001
LOCUS
DEFINITION Mus musculus putative retinoic acid-regulated protein mRNA,
complete cds.
ACCESSION AF361644
VERSION AF361644.1 GI:14028766
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1909)
REFERENCE Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
AUTHORS Trapping and characterization of novel retinoic acid response
TITLE elements
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1909)
AUTHORS Glozak, M.A., Li, Y., Reuille, R. and Rogers, M.B.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2001) Biology, University of South Florida, 4202
E. Fowler Avenue, Tampa, FL 33620, USA
Location/Qualifiers
1. .1909
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="P19; derived from matings between C3H/He

females and male animals carrying an X chromosome derived
from a feral mouse (McBourney & Rogers, Dev. Biol., 89,
503)"
/cell_type="embryonal carcinoma"
/clone_lib="24 hr retinoic acid-treated P19 cDNA (Bouillet
et al., Dev. Biol., 170, 420)"
89. .919
CDS
/codon_start=1
/product="putative retinoic acid-regulated protein"
/protein_id="AAK52494.1"
/db_xref="GI:14028767"

/translation="MESSELAAPRCCLFLLPRLPAPKLGSPAGAEETDWR
LPSCVEVKYAVELKSAFETGKTKEVIDTGYGLDGKSGVYKTSDLRIEVTET
ICKRLDYSLHKERTGNSRFAGKMSSETFETLHNLVHKGKVMVDIPYELNMTSAEVA
DLKKQCDLVVEEFEEVDWYRNHQQEEDLTFELCANHLVKGKDTSCLAERWSGKGGDI
ASLGKSKKKRSGVKGSSSGSKORKELGGLGEDANAEEREGVQKASPLPHSPDDEL
"

BASE COUNT 424 a 536 c 558 g 386 t 5 others
ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1909;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaagt 25
|||||
Db 216 TCGATTGCCAGCAAAATGCGAAGT 240

RESULT 3
AC074220/c 89855 bp DNA linear HTG 25-JAN-2002
LOCUS
DEFINITION Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 38 unordered pieces.

ACCESSION AC074220
VERSION AC074220.7 GI:13876390
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 89855)
REFERENCE Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
AUTHORS Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
TITLE High Throughput Mouse Sequencing
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 89855)
AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,
Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Apr 28, 2001 this sequence version replaced gi:11276104.
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous>
e.html
Contact: gntm@apecod.bwh.harvard.edu
-----Summary Statistics
Center project name: ADY
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 79355 at least Q20
*Consensus quality: 75365 at least Q30
*Consensus quality: 69022 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 89115 - sum-of-contigs

Quality coverage: agarose-FP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

7963	7962: contig of 7962 bp in length
7983	7982: gap of unknown length
13923	13922: contig of 5942 bp in length
13945	13944: gap of unknown length
18991	18990: contig of 5046 bp in length
19011	19010: gap of unknown length
22502	22501: contig of 3491 bp in length
22522	22521: gap of unknown length
22532	22531: contig of 4556 bp in length
27078	27077: gap of unknown length
27098	27097: contig of 1722 bp in length
28820	28819: gap of unknown length
28840	28839: contig of 3385 bp in length
31225	31224: gap of unknown length
31245	31244: gap of unknown length
35094	35093: contig of 3849 bp in length
35114	35113: gap of unknown length
38612	38611: contig of 3498 bp in length
38632	38631: gap of unknown length
42262	42261: contig of 3630 bp in length
42282	42281: gap of unknown length
44131	44130: contig of 1849 bp in length
44151	44150: gap of unknown length
47060	47059: contig of 2909 bp in length
47080	47079: gap of unknown length
49497	49496: contig of 2417 bp in length
49517	49516: gap of unknown length
52337	52336: contig of 2720 bp in length
52357	52356: gap of unknown length
54271	54270: contig of 2014 bp in length
54291	54290: gap of unknown length
56372	56291: contig of 1981 bp in length
56392	56291: gap of unknown length
58850	58849: contig of 2558 bp in length
58870	58869: gap of unknown length
60822	60821: contig of 1752 bp in length
60842	60841: gap of unknown length
62256	62255: contig of 1634 bp in length
62276	62275: gap of unknown length
63086	63085: contig of 810 bp in length
63106	63105: gap of unknown length
63441	63440: contig of 2335 bp in length
63461	63460: gap of unknown length
67630	67629: contig of 2169 bp in length
67650	67649: gap of unknown length
69179	69198: contig of 1529 bp in length
69199	69198: gap of unknown length
70096	70095: contig of 897 bp in length
70116	70115: gap of unknown length
71822	71821: contig of 1706 bp in length
71842	71841: gap of unknown length
73507	73506: contig of 1665 bp in length
73527	73526: gap of unknown length
74750	74749: contig of 1223 bp in length
74770	74769: gap of unknown length
75964	75963: contig of 1194 bp in length
75984	75983: gap of unknown length
77906	77905: contig of 1932 bp in length
77926	77925: gap of unknown length
79237	79236: contig of 1311 bp in length
79257	79256: gap of unknown length
81333	81332: contig of 2076 bp in length
81352	81351: gap of unknown length

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

Location/Qualifiers

```
1. .89855
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-76I16"
/sex="male"
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc feature
misc feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature 73527..74749
/note="assembly_name:Contig119"
74770..75963
/note="assembly_name:Contig118"
75984..77905
/note="assembly_name:Contig117"
77926..79236
/note="assembly_name:Contig116"
79257..81332
/note="assembly_name:Contig115"
81353..83169
/note="assembly_name:Contig114"
83190..84096

Query Match 96.0%; Score 24; DB 2; Length 89855;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgattgccagcaaatgcgaag 24
Db 5896 TCGATTGCCAGCAAAATGCGAAG 5873

RESULT 4
LOCUS AX079702 510 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 446 from Patent WO0107611.
ACCESSION AX079702
VERSION AX079702.1 GI:13159271
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 510)
Baker, K.P., Goddard, A. and Wood, W.I.
TITLE Human polypeptides and methods for the use thereof
JOURNAL Patent: WO 0107611-A 446 01-FEB-2001;
Genentech, Inc. (US)

FEATURES
source
1..510
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 122 a 114 c 168 g 104 t
ORIGIN

Query Match 85.6%; Score 21.4; DB 6; Length 510;
Best Local Similarity 95.7%; Pred. No. 3.2;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaag 25
Db 206 CGCTTGCCAGCAAAATGCGAAGT 228

RESULT 5
LOCUS AR142811 744 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 3 from patent US 6204013.
ACCESSION AR142811
VERSION AR142811.1 GI:15104097
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 744)
AUTHORS Khodadoust, M.M.
TITLE MSP-5 nucleic acid molecules and uses therefor
JOURNAL Patent: US 6204013-A 3 20-MAR-2001;
LOCATION/Qualifiers
FEATURES 1..744
source /organism="unknown"

BASE COUNT 208 a 143 c 254 g 139 t
ORIGIN

Query Match 85.6%; Score 21.4; DB 6; Length 744;
Best Local Similarity 95.7%; Pred. No. 3.4;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaag 25
Db 97 CGCTTGCCAGCAAAATGCGAAGT 119

RESULT 6
LOCUS AX136616 746 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 538 from Patent EP1067182.
ACCESSION AX136616
VERSION AX136616.1 GI:14273020
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 746)
Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and
Hayashi, K.

TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 538 10-JAN-2001;
Helix Research Institute (JP)

FEATURES
Location/Qualifiers
source
1..746
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 201 a 142 c 251 g 149 t
ORIGIN

Query Match 85.6%; Score 21.4; DB 6; Length 746;
Best Local Similarity 95.7%; Pred. No. 3.4;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 cgattgccagcaaatgcgaag 25
Db 154 CGCTTGCCAGCAAAATGCGAAGT 176

RESULT 7
LOCUS BD006701 747 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel polypeptide.
ACCESSION BD006701
VERSION BD006701.1 GI:18635072
KEYWORDS JP 2001029090-A/4.
SOURCE Homo sapiens.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 747)
AUTHORS Ito, Y., Mogi, S., Tanaka, H., Okubo, S. and Ogi, K.

TITLE Novel polypeptide
JOURNAL Patent: JP 2001029090-A 4 06-FEB-2001;
TAKEDA CHEMICAL INDUSTRIES LTD

COMMENT OS Homo sapiens (human)
PN JP 2001029090-A/4
PD 06-FEB-2001
PF 19-MAY-2000 JP 2000147530

PR PI YASUAKI ITO, SHINTACHI MOGI, HIDEYUKI TANAKA, SHOICHI OKUBO, PI
KAZUHIRO OGI
PC C12N15/09, A61K38/00, A61K39/395, A61K45/00, A61K48/00,
PC A61P1/00,
PC A61P5/00, A61P9/00, A61P11/00, A61P25/00, A61P31/00, PC


```

A61P35/00,
PC A61P37/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,G01N33/50,
PC G01N33/566//
PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/08,(C12P21/02, PC
C12R1:91),
PC C12N5/10,C12R1:91),(C12P21/08,C12R1:91),C12N15/00,A61K37/02,
PC C12N5/00,
CC (C12N5/00,C12R1:91)
FH Key Location/Qualifiers
FT source 1..747
FT /organism="Homo sapiens (human)".
FEATURES
    source
        1..747
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT 209 a 143 c 255 g 140 t
ORIGIN
Query Match 85.6%; Score 21.4; DB 6; Length 747;
Best Local Similarity 95.7%; Pred. No. 3.4;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 cgattgccagcaaatgcgaagt 25
|||
Db 97 CGCTTGCCAGCAAAATGCGAAGT 119

RESULT 8
BC019903 1413 bp mRNA linear PRI 22-JAN-2002
LOCUS Homo sapiens, clone IMAGE:4991480, mRNA, partial cds.
DEFINITION BC019903
ACCESSION BC019903
VERSION BC019903.1 GI:18043798
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1413)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amgbcn.tmc.edu
Gunatetne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 42 Row: j Column: 23.
    Location/Qualifiers
        1..1413
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4991480"
            /tissue_type="placenta, choriocarcinoma"
            /clone_id="NIH_MGC_10"
            /lab_host="DH10B"

```

```

/note="Vector: PCMV-SPORT6"
<1..774
/codon_start=1
/product="Unknown (protein for IMAGE:4991480)"
/protein_id="AAH19903.1"
/db_xref="GI:18043798"
/translation="OAVTCRRRVMPVRLGILLFLFLAVHEAWAGMKEDDDTERLPKCEVCKLL
LPSKCEVCKLLSTELQALSRYSREVLGQVLDTKRRRHVPYSVSETRLEALE
NLCEIRLDYSVHAERKSLRYAKGOSQTMATLKGVLQKGVKVDLGIPLLEWDEPSVEY
TYLKKCEETMLEEFIVGDWYFHHEQPLQNFCEGHVLPAAETACIQETWTGKEITDGEET
DGEETGEEEEEDEEGGDKMTKTSHPKLDREDL"
BASE COUNT 378 a 301 c 413 g 321 t
ORIGIN
Query Match 85.6%; Score 21.4; DB 9; Length 1413;
Best Local Similarity 95.7%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 cgattgccagcaaatgcgaagt 25
|||
Db 130 CGCTTGCCAGCAAAATGCGAAGT 152

RESULT 9
AX136423 1419 bp DNA linear PAT 30-MAY-2001
LOCUS AX136423
DEFINITION Sequence 345 from Patent EP1067182.
ACCESSION AX136423
VERSION AX136423.1 GI:14272827
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1419)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 345 10-JAN-2001;
Helix Research Institute (JPI)
FEATURES
    source
        1..1419
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="unnamed protein product"
            /codon_start=1
            /db_xref="GI:14272828"
            /translation="MGPVRLGILLFLFLAVHEAWAGMKEDDDTERLPKCEVCKLL
            STELOAELSRYSREVLGQVLDTKRRRHVPYSVSETRLEALENLCEIRLDYSV
            HAERKSLRYAKGOSQTMATLKGVLQKGVKVDLGIPLLEWDEPSVEYTKQCEETL
            EFEDIVGDWYFHHEQPLQNFCEGHVLPAAETACIQETWTGKEITDGEETGEE
            QEEEEEEDEEGGDKMTKTSHPKLDREDL"
BASE COUNT 363 a 307 c 422 g 327 t
ORIGIN
Query Match 85.6%; Score 21.4; DB 6; Length 1419;
Best Local Similarity 95.7%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 cgattgccagcaaatgcgaagt 25
|||
Db 154 CGCTTGCCAGCAAAATGCGAAGT 176

RESULT 10
AR142810 1512 bp DNA linear PAT 08-AUG-2001
LOCUS AR142810
DEFINITION Sequence 1 from patent US 6204013.
ACCESSION AR142810

```

5

JOURNAL

Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:14647315.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0506M12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgrl.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgrl.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-161A2, 2000 bp overlap; the clone sequenced to the right is RP13-741A20. Actual end is at base position 132764 of RP11-506M12.

Data from AC093619 was used to finish this clone, AC073842.

Polymorphisms have been identified between AC005071 and AC073842.

FEATURES

source

Location/Qualifiers

1..132764
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-506M12"
/clone_lib="RPCI-11"

misc_feature

1..397
/note="similar to EST AA009576 (NID:gl1470735) ze82d10.rl"

misc_feature

1..397
/note="similar to EST AW302599 (NID:g6712279) xw07b03.xl"

misc_feature

1..67
/note="similar to EST BE729033 (NID:gl0143025)"

misc_feature

1..67
/note="similar to EST BE666576 (NID:gl0027167)"

misc_feature

1..397
/note="similar to EST BF026637 (NID:gl0734349)"

repeat_region

462..513
/note="similar to EST AA009577 (NID:gl1470736) ze82d10.sl"

/rpt_family="L2"
572..673
/note="similar to EST BE729033 (NID:gl0143025)"
572..673
/note="similar to EST BF026637 (NID:gl0734349)"
727..1020
/note="similar to EST AA077273 (NID:gl1836747)"
881..1191
/rpt_family="Alu"
1165..1206
/rpt_family="(TAAA)n"
1214..1520
/rpt_family="Alu"
1495..1522
/rpt_family="(GAAA)n"
1603..1723
/rpt_family="L2"
2050..2494
/note="similar to EST AW953106 (NID:gl142789)"
2090..2242
/note="similar to EST T08560 (NID:g389588)"
2104..2240
/note="similar to EST BE666576 (NID:gl0027167)"
2104..2240
/note="similar to EST BF026637 (NID:gl0734349)"
2104..2240
/note="similar to EST BE729033 (NID:gl0143025)"
2107..2493
/note="similar to EST R85198 (NID:g943604) y041a11.rl"
2140..2240
/note="similar to EST AW258515 (NID:g6631496) uq36b09.yl"
2298..3060
/note="similar to EST AV723000 (NID:gl0826061)"
2347..2832
/note="similar to EST H18846 (NID:g885086) ym45e12.rl"
2347..2522
/note="similar to EST BE534256 (NID:g9762901)"
2354..2462
/note="similar to EST BE666576 (NID:gl0027167)"
2354..2522
/note="similar to EST BE729033 (NID:gl0143025)"
2354..2522
/note="similar to EST AW258515 (NID:g6631496) uq36b09.yl"
2354..2522
/note="similar to EST BF026637 (NID:gl0734349)"
2359..2522
/note="similar to EST BE727269 (NID:gl0141362)"
2407..2522
/note="similar to EST BE664898 (NID:gl0024867)"
2478..2837
/note="similar to EST BF510418 (NID:gl1593716)"
2533..3040
/note="similar to EST H17220 (NID:g883460) ym42c05.rl"
2535..3115
/note="similar to EST T75205 (NID:g691967) yc86e12.rl"
2653..2725
/note="similar to EST BE729033 (NID:gl0143025)"
2653..2745
/note="similar to EST AW258515 (NID:g6631496) uq36b09.yl"
2653..2745
/note="similar to EST BE664898 (NID:gl0024867)"
2653..2725
/note="similar to EST BE727269 (NID:gl0141362)"
2653..2725
/note="similar to EST BF026637 (NID:gl0734349)"
2653..2745
/note="similar to EST BE534256 (NID:g9762901)"
2742..2987
/note="similar to EST Z42144 (NID:g564385)"
2800..3015
/note="similar to EST BI037302 (NID:gl14443928)"
2828..2908
/note="similar to EST BE664898 (NID:gl0024867)"

```
misc_feature 2828. .2908
/note="similar to EST AW258515 (NID:g6631496) uq36b09.y1"
misc_feature 2828. .2908
/note="similar to EST BE534256 (NID:g9762901)"
misc_feature 2863. .3309
/note="similar to EST R54265 (NID:g816167) yg74h03.rl"
misc_feature 2893. .3290
/note="similar to EST H14061 (NID:g878909) ym63a03.s2"
misc_feature 2905. .3554
/note="similar to EST BF921437 (NID:g12317325)"
misc_feature 3343. .3430
/note="similar to EST BE534256 (NID:g9762901)"
misc_feature 3343. .3430
/note="similar to EST BE664898 (NID:g10024867)"
misc_feature 3343. .3428
/note="similar to EST AW258515 (NID:g6631496) uq36b09.y1"
misc_feature 3434. .3950
/note="similar to EST AAL69431 (NID:g1748398) zo84g01.rl"
misc_feature 3487. .4034
/note="similar to EST BF939888 (NID:g12357208)"
misc_feature 3506. .3669
/note="similar to EST BE073223 (NID:g8420006)"
misc_feature 3534. .3701
/note="similar to EST BE664898 (NID:g10024867)"
misc_feature 3549. .4041
/note="similar to EST H17115 (NID:g883355) ym42c05.sl"
misc_feature 3681. .4041

Query Match 81.68; Score 20.4; DB 9; Length 132764;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 cgattgccagcaaatgcgaag 24
||| ||||| ||||| ||||| |||||
Db 84853 GCCTGCCAGCAATGCGAAG 84832

RESULT 13
AC092483/3
LOCUS AC092483 167854 bp DNA linear HTG 21-AUG-2001
DEFINITION Homo sapiens chromosome UNK clone CTD-2188L12, *** SEQUENCING IN
PROGRESS ***, 44 unordered pieces.
ACCESSION AC092483
VERSION AC092483.2 GI:15217361
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167854)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 167854)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 21, 2001 this sequence version replaced gi:14647307.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.MS2188L12
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing method: plasmid; 8%
Chemistry: Dye-terminator ET; 3% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads
Assembly program: Phrap; version 0.990319
```

```
Consensus quality: 115486 bases at least Q40
Consensus quality: 134295 bases at least Q30
Consensus quality: 146715 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 163554; sum-of-contigs
Quality coverage: 2.74 in Q20 bases; agarose-fp
Quality coverage: 2.79 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1021: contig of 1021 bp in length
* 1121: gap of unknown length
* 1122: contig of 1603 bp in length
* 2724: gap of unknown length
* 2824: gap of unknown length
* 3877: contig of 1053 bp in length
* 3878: gap of unknown length
* 3977: gap of unknown length
* 5593: contig of 1616 bp in length
* 5594: gap of unknown length
* 5694: contig of 2721 bp in length
* 8414: contig of unknown length
* 8514: gap of unknown length
* 8515: contig of 1204 bp in length
* 9718: gap of unknown length
* 9818: gap of unknown length
* 9819: contig of 2697 bp in length
* 12515: gap of unknown length
* 12616: contig of 1775 bp in length
* 14391: gap of unknown length
* 14491: contig of 2870 bp in length
* 17361: gap of unknown length
* 17461: contig of 2071 bp in length
* 19532: gap of unknown length
* 19632: contig of 2500 bp in length
* 22132: gap of unknown length
* 22332: contig of 1462 bp in length
* 23694: gap of unknown length
* 23794: gap of unknown length
* 26182: contig of 2389 bp in length
* 26282: gap of unknown length
* 26283: contig of 1867 bp in length
* 28150: gap of unknown length
* 28249: gap of unknown length
* 28250: contig of 1767 bp in length
* 30017: gap of unknown length
* 30117: contig of 1781 bp in length
* 31897: gap of unknown length
* 31997: contig of 2155 bp in length
* 34152: gap of unknown length
* 34252: gap of unknown length
* 36342: contig of 2090 bp in length
* 36442: gap of unknown length
* 40089: contig of 3647 bp in length
* 40189: gap of unknown length
* 42652: contig of 2463 bp in length
* 42752: gap of unknown length
* 45366: contig of 2614 bp in length
* 45367: gap of unknown length
* 45467: contig of 2564 bp in length
* 48031: gap of unknown length
* 48130: contig of 3255 bp in length
* 51385: gap of unknown length
* 51486: contig of 3522 bp in length
* 55007: gap of unknown length
* 55107: contig of 4321 bp in length
* 59428: gap of unknown length
* 59528: contig of 4375 bp in length
* 63903: gap of unknown length
* 64003: contig of 4554 bp in length
* 68557: gap of unknown length
* 68558: contig of 5391 bp in length
* 74048: gap of unknown length
* 74148: contig of 4600 bp in length
* 78748: contig of 4600 bp in length
```

```

* 78749 78848: gap of unknown length
* 78849 83676: contig of 4828 bp in length
* 83677 83776: gap of unknown length
* 83777 87313: contig of 3537 bp in length
* 87314 87413: gap of unknown length
* 87414 91962: contig of 4549 bp in length
* 91963 92062: gap of unknown length
* 92063 97831: contig of 5769 bp in length
* 97832 97931: gap of unknown length
* 97932 102267: contig of 4336 bp in length
* 102268 102367: gap of unknown length
* 102368 107501: contig of 5134 bp in length
* 107502 107601: gap of unknown length
* 107602 112585: contig of 4984 bp in length
* 112586 112685: gap of unknown length
* 112686 11771: contig of 5086 bp in length
* 11772 11781: gap of unknown length
* 11782 123681: contig of 5810 bp in length
* 123682 123781: gap of unknown length
* 123782 129137: contig of 5356 bp in length
* 129138 129237: gap of unknown length
* 129238 135973: contig of 6736 bp in length
* 135974 136073: gap of unknown length
* 136074 144630: contig of 8557 bp in length
* 144631 144730: gap of unknown length
* 144731 152830: contig of 8100 bp in length
* 152831 152930: gap of unknown length
* 152931 161016: contig of 8086 bp in length
* 161017 161116: gap of unknown length
* 161117 167854: contig of 6738 bp in length.

FEATURES
    source
        Location/Qualifiers
            1..167854
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="UNK"
                /clone="CTD-2188L12"
            1..1021
                /note="assembly_name:Contig20"
            1122..2724
                /note="assembly_name:Contig21"
            2825..3877
                /note="assembly_name:Contig22"
            3978..5593
                /note="assembly_name:Contig23"
            5694..8414
                /note="assembly_name:Contig24"
            8515..9718
                /note="assembly_name:Contig25"
            9819..12515
                /note="assembly_name:Contig26"
            12616..14390
                /note="assembly_name:Contig27"
            14491..17360
                /note="assembly_name:Contig28"
            17461..19531
                /note="assembly_name:Contig29"
            19632..22131
                /note="assembly_name:Contig30"
            22232..23693
                /note="assembly_name:Contig31"
            23794..26182
                /note="assembly_name:Contig32"
            26283..28149
                /note="assembly_name:Contig33"
            28250..30016
                /note="assembly_name:Contig34"
            30117..31897
                /note="assembly_name:Contig35"
            31998..34152
                /note="assembly_name:Contig36"
            34253..36342
                /note="assembly_name:Contig37"
            36443..40089

```

```

misc_feature /note="assembly_name:Contig38"
40190..42652
misc_feature /note="assembly_name:Contig39"
42753..45366
misc_feature /note="assembly_name:Contig40"
45467..48030
misc_feature /note="assembly_name:Contig41"
48131..51385
misc_feature /note="assembly_name:Contig42"
51486..55007
misc_feature /note="assembly_name:Contig43"
55108..59428
misc_feature /note="assembly_name:Contig44"
59529..63903
misc_feature /note="assembly_name:Contig45"
64004..68557
misc_feature /note="assembly_name:Contig46"
68658..74048
misc_feature /note="assembly_name:Contig47"
74149..78748

Query Match      81.6%   Score 20.4; DB 2; Length 167854;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 cgattgccagcaaatgcgaag 24
|||
Db 41488 CGTTGCCAGCAAAATGCGAAG 41467

RESULT 14
AC087135
LOCUS AC087135 209958 bp DNA linear HTG 09-DEC-2000
DEFINITION Mus musculus clone RP23-147H4, WORKING DRAFT SEQUENCE, 31 unordered
pieces.
ACCESSION AC087135
VERSION AC087135.1 GI:11610861
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 209958)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1793492
Center clone name: RPCI-23_147H4
-----
Summary Statistics
Consensus quality: 180916 bases at least Q40
Consensus quality: 198060 bases at least Q30
Consensus quality: 201064 bases at least Q20
Estimated insert size: 204000; agarose-fp estimation
Estimated insert size: 206958; sum-of-contigs estimation
Quality coverage: 8.82 in Q20 bases; agarose-fp estimation
Quality coverage: 8.69 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 23, 2002, 23:05:31 ; Search time 4203.75 Seconds
(without alignments)
5904.466 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgagggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1528	83.1	1553	11	AK005532
2	1345.6	73.2	1586	11	AK010702 Mus muscu
3	761.2	41.4	766	10	BG974240 602843972
4	743.8	40.4	793	10	BI104487 602890091
5	722.2	39.3	985	10	BF785328 602111557
6	713	38.8	737	10	BG873094 602794679
7	711.4	38.7	836	10	BI078951 602873203
8	677.8	36.9	942	10	BF385997 602046176
9	646.8	35.2	698	10	BI110829 602895470
10	641.8	34.9	779	10	BI688404 60316338
11	638.2	34.7	645	9	BA498921
12	637.8	34.7	641	9	BB637089
13	632.8	34.4	644	9	BB190512
14	628.4	34.2	656	10	BI082275
15	626.4	34.1	651	9	BB385488
16	626	34.0	626	10	BM232538
17	626	34.0	627	9	BB531047

18	620.8	33.8	776	10	BG864846
19	619.6	33.7	925	9	AL520942
20	618.4	33.6	1051	10	BM471899
21	613.6	33.4	632	9	BB558801
22	611.8	33.3	656	9	BB504341
23	608	33.1	643	9	BB61603
24	607.4	33.0	625	9	BB633565
25	606.8	33.0	790	10	BF301937
26	606.4	33.0	741	10	BI081472
27	605.6	32.9	829	10	BI833903
28	603.2	32.8	608	9	BB605812
29	599.6	32.6	815	9	AL559536
30	594.2	32.3	665	9	BB218998
31	591.4	32.2	838	10	BF334535
32	586	31.9	916	10	BE790646
33	583	31.7	812	10	BI918555
34	574.6	31.2	808	10	BG686355
35	573.8	31.2	893	10	BG685190
36	571.4	31.1	573	9	AA260027
37	568.2	30.9	713	10	BM042021
38	566.2	30.8	1020	10	BF975546
39	553	30.1	553	10	BM223931
40	547	29.7	666	10	BG681796
41	545.6	29.7	733	10	BI764820
42	541	29.4	623	9	BB661026
43	540.6	29.4	669	10	BG866446
44	537.4	29.2	1003	10	BE250104
45	534.8	29.1	649	9	BB652011

ALIGNMENTS

RESULT 1

AK005532	LOCUS	1553 bp	mRNA	linear	HTC 19-JAN-2002
DEFINITION	Mus musculus adult female placenta cDNA, RIKEN full-length enriched library, clone:1600025D17:homolog to CTG4A, full insert sequence.				
ACCESSION	AK005532.1 GI:12838152				
VERSION	HTC; CAP trapper.				
KEYWORDS	Mus musculus (strain:C57BL/6J) adult female placenta cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library				
SOURCE	clone:1600025D17.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	1 (sites)				
TITLE	Carninci, P. and Hayashizaki, Y.				
JOURNAL	High-efficiency full-length cDNA cloning				
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	2 (sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				

Db 275 GGTGTGTATGGATATCCCOCTATGAGCTGTGGAACGAGACCTCAGCAGAGGTGGCTGACT 334
Qy 540 caaagaacagtgatgactgctgctggtgaaagagtttgaagaggtgattgaggaactggtacag 599
Db 335 CAAGAAGACAGTGTGACGTGCTGCTGGTGGAGAGATTTCGAAGAGTGTATGAGGACTGTTACAG 394
Qy 600 gaaccacacagaggaagacacctgactgaattctctctgtgtcccaaccacgctgctgaaagggaaa 659
Db 395 GAACCACACAGGAGGAAGACCTGACTGAATTCCTCTGTGCCAACACACAGTGTCTGAAGGGAAA 454
Qy 660 ggacacagagtgccttagacagagcgtgctctgcaagaaggggacacatagcctccctggg 719
Db 455 GGACACGAGTGTCCCTAGCAGAGCGGTGTCTGGCAAGAAAGGGGACATAGCCTCCCTGGG 514
Qy 720 aggaagaatactcaagaagaagcagcgagtgcaaggcctctccagtgccagcagcaaa 779
Db 515 AGGGAAGAAATCCAGAGAGAGCGCACGCGAGTCAAGGGCTCTCCAGTGGCAGCAGCAA 574
Qy 780 gcagagggaagaactgggggctgggggagagatcccaacgcccagggagggaggggtgt 839
Db 575 GCAGAGGAAGAACTGGGGGGCTGGGGGAGGATGCCAACGCCGAGGAGGAGGGTGT 634
Qy 840 qcaaaagcatcg -cccccccaacacagccccccctctgtagctgagccagcttagtg 898
Db 635 GCAGAAGGCATCGCCCCCTCCACACAGCCCCCTGTAGAGCTGTGAGCCAGCTTAGTG 694
Qy 899 tccctgaatcaagacccttgacttcagagcttggaacacgcacagcgagcgagcgag 958
Db 695 TCCTTGAATCAAGACCCCTGACTTCAGAGCTTGGGACACGCACACGCGCAGCGCAG 754
Qy 959 ctccgcaaggaagactgctgtctccagcatcaggtctctccctgtgctgctccctttcc 1018
Db 755 CTCACGACAGAGAGCTGTGTCTCCAGCATCAGGTCTCTCTCTTGGCTGTGCCCCCTTCC 814
Qy 1019 tccctctaaacaacagcaagagtgaaagatctgggtgctggtgagagcgaccacccaaa 1078
Db 815 TTCCCTTGAACAACAGCAAGAGGTGGAAGGATCTGGGTGCTGGAGACGCGACCCCAAA 874
Qy 1079 gggaagagagagagagagagagagctctctttctacacagtcctccctcagagctcc 1138
Db 875 GGAAGAGAGGAGGAGGAGCAGAGGAGCAGCTCTCTTTCTACACAGTCCCCCTCAGAGTCC 934
Qy 1139 ggggtccaccacagatccccagctgagatccaggtctctgacatggagctgaagagca 1198
Db 935 GGGGTCCACCCAGCATCCCGAGCTGAGATCCAGGCTCCTTGACATGGAAGCTGAAGAGCA 994
Qy 1199 tgaggcacataagatgctcaccagcgcccccttcagccaggaagagactccgtgcagctcc 1258
Db 995 TGAGGCACATAGATGCTCACCAGCGCCCCCTTCAGCCAGGAAGGACTCCGTCGACCTC 1054
Qy 1259 agcagcaggcctgctctcttcctccacaagcatctcttctgctggtctgctgctgat 1318
Db 1055 AGCAGCCAGGCTGCTCTCTCTCCACCAAGCATCTCTTCTGTGTGCTCTGTCGAT 1114
Qy 1319 ggttaaticgagaactccagagcaaaactcgggtgtgcacaaagggctgagaccagca 1378
Db 1115 GGTAAATTCGAGAACTTCAGGACAAACTCGGGGTGTGGCACAAGAGGGCTGGACGCCAGA 1174
Qy 1379 gscagagccacgcagagactgcagagagggcacctgacctaaccctccctggaaagccaa 1438
Db 1175 GCCAGAGCCAGCCAGAGACTGCAGAGGGGACCTGACCTAACCCCTCGAAGGCCAA 1234
Qy 1439 tctgcagttcccggttccaccacatctctctgagagcctcatgctctgcccagccctt 1498
Db 1235 TCTGCAGTTCCCGTGTCCACCCACTCTCTCTGAGGACGCTCATGCTGTGCCAGCCCTT 1294
Qy 1499 ctcccagggctaccagagtaaacacttttggcccttgggttgggttctggtggtcctcat 1558
Db 1295 CTCACAGGGCTACCAAGATTAACACCTTTTGGCCCTTTGCGTTGTTGCTTGGGTCTCAT 1354
Qy 1559 cagcctccagagtgctccctcatcatcttttttccctttgtcccccataccaggggct 1618
Db 1355 CAGCCTCCAGAGTGTCCCTCATCATCTTTTTCGCCCTTGTGCCCAATCCACAGGGCT 1414

QY 1619 ggaagggccatcaccatcattgaggttaacctgtcagttactaggaggtgctgggagcg 1678
Db 1415 GGAAGGCCATCACCATCATTTGGAGGCTTAACCTGTCTAGTTACTAGGAGGTGCTGGGAGCG 1474
QY 1679 cccggggttggttgggtaatacactcaactggtcctcagccttctaaactgcagccct 1738
Db 1475 CCGGGGTTGGTTGGGTAATCACTACTGGCTCTCAACCTTCTTAACACTGCGAGCCCT 1534
QY 1739 taatacagttcctctgttgggtgagctccacgccccccacacacacacat 1790
Db 1535 TAATACAGTTCCTTCTGTGTGTGACTCCAGCGCCCCCACACACACACAT 1586
RESULT 3
LOCUS BG974240 766 bp mRNA linear EST 12-JUN-2001
DEFINITION 602843972F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979944 5',
mRNA sequence.
ACCESSION BG974240
VERSION BG974240.1 GI:14361877
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 766)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999).
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10978 row: o column: 17
High quality sequence stop: 761.
FEATURES
Location/Qualifiers
1..766
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4979944"
/clone_lib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo df.
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
Library."

BASE COUNT 185 a 228 c 236 g 117 t
ORIGIN

Query Match 41.4%; Score 761.2; DB 10; Length 766;
Best Local Similarity 99.6%; Pred. No. 3.6e-131;
Matches 763; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 658 aaggacacagtgctcctagcagcggtgtctgccaagaaggagggacatagcctccctg 717
Db 1 AAGGACACAGTGTCTCTAGCAGAGCGGTGTCTGGCAAGAGGGGACATAGCTCCCTG 60
QY 718 ggagggaagaataccaagaagacgacgagtgctcaagggtcctccagtggcagcagc 777
Db 61 GGAGGGGAAGAAATCCAAAGAAGACGCGCAGCGGAGTCAAGGGCTCTCCAGTGCAGCAGC 120

ORIGIN

Query Match 38.8%; Score 713; DB 10; Length 737;
Best Local Similarity 98.6%; Pred. No. 3e-122;
Matches 719; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1096 cagaaggagctctctttctacacagtcctccctcagctcagctcgggtcccccagcgc 1155
|||
Db 6 CACGGCTCGCTCTCTTTCTACACAGTCCCTCTACAGAGTCCGGGGTCCACCCAGCATC 65
Qy 1156 cccaggctgagatccaggtcctctgacatggaagctgaagacatgagggcacataagatgc 1215
|||
Db 66 CCCAGGCTCAGATCCAGGCTCTGTACATGGAAGCTGAAGAGCATGAGGCACATAGATGC 125
Qy 1216 tcaccagagcccccttcagccaggaaggaactccgtgagcctcagcagcagggcctgct 1275
|||
Db 126 TCACAGGCGCCCCCTTCAGCCAGGAAGGACTCCGCTGCAGCTCAGACAGCCAGGCTGCCT 185
Qy 1276 ctctctccacacagcattctctctgctggtccctgtcgatggtataatcgaagactt 1335
|||
Db 186 CTCTCTCCACCAAGCATTCCTTCTGCTGTGCTCTGTGCGATGTTAAATTCGAGAACTT 245
Qy 1336 ccaggacaaactcgggtgtgacaaaggggtgagccagagccagagccagccagcaga 1395
|||
Db 246 CCAGGACAAACTCGGGTGTGGCACAAGGGGTGGACGCCAGAGCCAGCCAGCCAGCA 305
Qy 1396 gactgacagagggcacctgacctaaccctggaaagccaatctgagttcccgctgc 1455
|||
Db 306 GACTGCAGAGAGGGCACCTGACCTTAACCCCTTGGGAAGCCAAATCTGCAGTTCCCGTGT 365
Qy 1456 caccactctctcagagagcctcctatctctgcccagcctctctccagggctaccaga 1515
|||
Db 366 CACCCACTCTCTCTGAGAGCGCTCATCTCTGCCCGCCCTCTCTCCAGGCTACCCAGA 425
Qy 1516 gtaaacacttttggccttctggttggcttctcagcctcagcagtgctcc 1575
|||
Db 426 GTAACACACTTTTGGCTTTTGGCTTTGGCTTCTGCTGCTCTCATCAGCCTCCAGAGTGC 485
Qy 1576 cctcatgactcttttgcctcttctcccaatcccaagggctggaagccataccatc 1635
|||
Db 486 CTTATCATCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTC 545
Qy 1636 attgaggcttaactgctcagtaggagtggtgagcggcgggtggttgg 1695
|||
Db 546 ATTGAGGCTTAACCTGTCACTTACTAGAGGTGCTGGAGCGCCGGGTGCTATGG 605
Qy 1696 gtaatacactggtctcagccttcttaacactgagcccttaatacagttcctctg 1755
|||
Db 606 GTAATCACTCACTGGCTCTCAGCCTTCTTAACACTGCAGCCCTTAATACAGTTCTCTG 665
Qy 1756 ttgtggtgactccacgccccacacacacacacacataaaattatttcgatgctgttctca 1815
|||
Db 666 TTGTGCTCACTCCCGCCCCCGCCAGACACACAGCATAAAATATTTCGATGCTGTTTCATA 725
Qy 1816 actgtaaaa 1824
|||
Db 726 ACTGTACAA 734

RESULT 7

B1078951
LOCUS 602873203F1 NCI_CGAP_Mam2 Mus musculus cdna clone IMAGE:5005033 5',
DEFINITION mRNA sequence.
ACCESSION B1078951
VERSION B1078951.1 GI:14497281
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 836)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11044 row: e column: 02
High quality sequence stop: 707.

FEATURES

source

1. 836
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5005033"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT

187 a 265 c 210 g 174 t

ORIGIN

Query Match 38.7%; Score 711.4; DB 10; Length 836;
Best Local Similarity 95.2%; Pred. No. 5.9e-122;
Matches 768; Conservative 0; Mismatches 31; Indels 8; Gaps 3;

Qy 1041 gtgaaagatctgggtgctggagagcgcccccacaaagggaagagagggagcagaa 1100
Db 1 GTGAGAGATCTGGGGTCTGGGAGAGCGCACCCCAAGGGAAGAGAGGAGCAGAA 60
Qy 1101 ggcagctctctttacacagtcctccctcacyagctccgggttccacccagcatcccccag 1160
|||
Db 61 GGCAGCTCTCTTTACACAGTCCCTCTACAGAGTCCGGGTCCAGCCAGCATCCCGAG 120
Qy 1161 gctagatccaggtcctgacatgaagctgaagcagcagcagcagcagcagcagcagcagc 1220
|||
Db 121 GCTGAGATCCAGGCTCTCTGACATGGAAGCTGAAGAGCATGAGGACATGAAGATGCTCACC 180
Qy 1221 agcggcccttcagccaggaagagcctcgtcagcctcagcagcagcagcagcagcagcagc 1280
|||
Db 181 AGCGCCCTTTCAGCCAGGAAGGACTCCGTGACGCTCAGCAGCCAGCCAGCCAGCCAGCTTCC 240
Qy 1281 ttccacaaagcattctcttctgctggttctgctgctgctgctgctgctgctgctgctgctgct 1340
|||
Db 241 TTCCACCAAGCATCTCTCTCTGCTGGTCTGCTGGATGGTAAATTCGAGAACTTCCAGG 300
Qy 1341 acaactcgggtgtgacacaaagggtgagcgcacagccagccagccagccagccagccagccagcc 1400
|||
Db 301 ACAACTCGGGGTGGCACAAGGGGTGGACGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCAGT 360
Qy 1401 cagagagggcacctgacctgaactaaccccccttgaaagccaatctgcagttccctgctccaccc 1460
|||
Db 361 CAGAGAGGCGACCTGACCTAACCCCTTGGGAAGCCAAATCTGCAGTTCCCGTGTCCACCC 420
Qy 1461 actcctcctgagagcagcctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1520
|||
Db 421 ACTCCTCTGAGAGCGGCTCATGCTCTGCCAGCCCTTCTCCAGGGCTACCCAGAGTAAA 480
Qy 1521 cacttttggccttctggttgggttctcctggttctcctcctcctcctcctcctcctcctcctcctca 1580
|||
Db 481 CACCTTTGGCTTTTCGCTTTGGTTCTCTGGGTCTCTATCAGCTCCAGAGTGTCCCTCTCA 540
Qy 1581 tcgatcttttttgcctttgttcccccacaaatccca-ggggctggaaggccatcaccatcattg 1639


```
Qy 367 aagagctcttgactacagcctgcacaaagagagagactggcagcaaacccggtttgccaag 426
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 360 AAGAGGGTTCTGGACTACAGCTGCACAAGGAGGAGGACTGCGCAGCAACCGGTTGGCCAAG 419
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 427 ggtatgtcggagacctttgagacgctgcacaacctagtcacaaagggttcaagtggtg 486
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 420 GGTATGTGGAGACCTTTGAGACGGCTGCACAACCTTA-TCCCAAAAGGGGTCAAGGTGGTG 478
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 487 atgatatccctctatgag-cfctgaaacagacacctcagacagtggtggtgacctcaagaa 545
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 479 ATGGATATCCCTATGAGCCCTGTGAACGACGACCTTCACAGAGGTGGCTGACCTCAAGAA 538
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 546 gcagtgtagctgctggtggaa-gagtttgaagaggtgattgaggaactgggtacaggaacc 604
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 539 GCAGTGTACGCTGCTGGTGAAGGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACC 598
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 605 accaggaagaagacctgactgaattcctctgtgcacaaccacgctgctgaaggaagagaca 664
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 599 ACCAGGAGGAAGACCTGACTGAATTCCTGTGTGCCAACACGCTTCTGAAGGGAAGGACA 658
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 665 cgagtgctagcagacggtggctgctgcaagaaggagacatagcctccctggagggga 724
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 659 CGAGTTGCTAGCAGAGCGGTGGCTGTGCAG-AGGGGACATAGCCTTCCTGGGAGGGGA 717
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Qy 725 agaaatccaaagaagcgagcgagtgcaagggtcctccagtggcagcagca 779
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 718 AGAATCCAAAGAAAGACGCGACGGAATCAA-GGCTCTCCATCGCAGCAGCA 771
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 11
BB498921
LOCUS BB498921 RIKEN full-length enriched, 0 day neonate kidney Mus
DEFINITION musculus cDNA clone D630017H14.3', mRNA sequence.
ACCESSION BB498921
VERSION BB498921.2 GI:16442133
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 645)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al., 2001)
Unpublished (2001)
On Jul 26, 2000 this sequence version replaced gi:9474466.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillar sequencer. Genome Res. .
```


Db	301	GCCTCATGCTGCGCCAGCCCTTCTCCAGAGGCTACACAGATTAACACCTTTTGGCCTTT	360
Qy	1536	cggtttggttcttgggtctcatcagctccagagtgccccctcatcagatctttttgccc	1595
Db	361	CGGTTTGTTGCTGGTCCCTCATCAGCTCCAGAGTGTCCTCCCTCATCATCTTTTGGCC	420
Qy	1596	tttgcccccaatccccagggggctggaaggccatcaccatcattggagcttaacctgtca	1655
Db	421	TTTGTCCCCCAATCCAGAGGCTGGAAGGCCATCACCATCATTTGAGGCTTAACCTGTCA	480
Qy	1656	gttactaggaggtactggagcgccggggtgtgttggggtaatacactcactggctctc	1715
Db	481	GTTACTAGGAGGTGCTGGGAGCGCCGGGTGTGTGGGTAAATCACTACTGGGCTCTC	540
Qy	1716	agccttcaactcgagcccttaatacacgttctctctgttggtagactccacgccc	1775
Db	541	AGCCTTCTAACACTGCAGCCCTTAATACAGTTCTTCTGTGTGGTGAGTACCCACGCCC	600
Qy	1776	ccacacacacacataaaattatttcgatgctgtttcataaactg	1819
Db	601	CCACACACACACCAATAAATATTTCGATGCTGTTTCATAACTG	644
RESULT	12		
BB637089			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			


```

Qy 1725 acactgcagcccttaacacagttctctgtgtgtgactccacagcccccacacaca 1784
|||||
Db 549 ACACAGCCCTTATATACAGTCTCTCTGTGTGTGACTCCACGCCCCACACACA 608
|||||
Qy 1785 caccataaaattattcgatgctgttctcaactgt 1820
|||||
Db 609 CACCATAAAATATTTGATGCTGTTTCATAACTGT 644
|||||

RESULT 14
BI082275
LOCUS 602877456f1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5009148 5',
DEFINITION mRNA sequence. linear EST 20-JUN-2001
ACCESSION BI082275
VERSION BI082275.1 GI:14500605
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 656)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1054 row: p column: 13
High quality sequence stop: 656.
FEATURES
Location/Qualifiers
1..656
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5009148"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 147 a 207 c 197 g 105 t
ORIGIN
Query Match 34.2%; Score 628.4; DB 10; Length 656;
Best Local Similarity 99.7%; Pred. No. 1.4e-106;
Matches 640; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 741 gcgcagcgagtcagggtctctccagtcgagcagcagcagaggaaggaactgggggg 800
|||||
Db 13 GCGCAGCGGAGTCGAGGCTCTCCAGTGGCAGCAGCAGCAGGAGGAAGTGGGGG 72
|||||
Qy 801 cctgggggagatgccaacgcgagagagagaggggtgtgagagagcagcagcctccc 860
|||||
Db 73 CTTGGGGGAGGATGCCAACGCCGAGGAGGAGGGTGTGAGAGGCGATGCCCTCC 132
|||||
Qy 861 acacagcccccctgatgagctgtgagccagcttagtctctgaatcaagaccctgac 920
|||||
Db 133 ACACAGCCCCCTGATGACCTGTGAGCCACGCTTAGTCTCTTGNATCAACACCCTGAC 192
|||||
Qy 921 ttccagagcttgggacacacagcagcagcagcagcagcagcagcagcagcagctgctgt 980
|||||

```

```

Db 193 TTACAGAGCTTGGGACACGCACACGCGCAGCGCAGCTCCACGAAGGACAGCTGCTGT 252
Qy 981 ccagcatcaggctcctcccttggctgtgcccccttcttcccttgaaacacagagag 1040
|||||
Db 253 CCACGATCAGGTCCTCCCTTGGCTGTGCCCCCTTTCCTTCCCTTGAACAACAGCAAGAG 312
|||||
Qy 1041 gtggaagagatcctgggtgctggagagcgcccccaaggaagagggagggagcagaa 1100
|||||
Db 313 GTGGAAGGATCTGGGGTCTGGGAGACGGCACCCCAAGGGAAGGAGGAGGAGCAGAA 372
|||||
Qy 1101 ggcagctctctttacacagctcccccttcacgagctccggggtccacccagcatccccc 1160
|||||
Db 373 GGCAGCTCTCTTTTACACAGTCCCTCTACAGAGCTCCGGGTCCACCCAGCATCCCCAG 432
|||||
Qy 1161 gctagatcaggctcctcgatgaaagctgaagacatgagcacataagatcactccc 1220
|||||
Db 433 GCTGAGATCCAGGCTCCTTGACATGAAGCTGAAGAGCATGAGGCACATAAAGATGCTCACC 492
|||||
Qy 1221 agcgcctccttcagccaggaaggaagctccgctgagcctcagcagcagcagcagcctgctctccc 1280
|||||
Db 493 AGCGCCCCCTTCAGCCAGGAGGACTCCGTGCAGCCTCAGCAGCCAGCGCTTCCTCTTCC 552
|||||
Qy 1281 ttccaccaagcattctcttctgctggtcctgtgctggtgtaaatcgagaactccagg 1340
|||||
Db 553 TTCCACCAAGCATTCTCTTCTGCTGGTGTGCGATGGTACATTTCGAGAAGCTTCCAGG 612
|||||
Qy 1341 acaaacctcgggtggtggcacaaaagggg-ctggagccagcagagcc 1381
|||||
Db 613 ACAAACTCGGGTGTGGCACAAAGGGGCTGGACGCCAGAGCC 654
|||||

RESULT 15
BI085488
LOCUS BB385488 RIKEN full-length enriched, 0 day neonate cerebellum Mus
DEFINITION musculus cDNA clone C230039G09 3', mRNA sequence.
ACCESSION BB385488
VERSION BB385488.2 GI:16408338
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 651)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
JOURNAL On Jul 13, 2000 this sequence version replaced gi:9108269.
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format

```

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES	SOURCE
1. High Accuracy: The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.	Model Performance Metrics
2. Scalability: The model is designed to handle large-scale data and complex tasks, making it suitable for enterprise-level applications.	System Architecture
3. Interpretability: The model's decisions are transparent and explainable, allowing users to understand the underlying reasoning.	Model Explainability
4. Robustness: The model is resilient to adversarial attacks and maintains performance in noisy or incomplete data environments.	Security and Reliability
5. Efficiency: The model is optimized for fast inference and low resource consumption, ensuring high performance in real-time applications.	Performance Optimization
6. Flexibility: The model can be adapted to different domains and tasks, providing a versatile solution for various use cases.	Domain Adaptability
7. Integration: The model seamlessly integrates with existing systems and workflows, facilitating easy adoption and deployment.	System Integration
8. Customization: The model can be tailored to meet specific requirements and preferences, offering a personalized experience.	Customization Options
9. Support: Comprehensive documentation and technical support are provided to assist users in maximizing the model's potential.	User Support and Training
10. Future Updates: The model is continuously updated with the latest research and advancements, ensuring it remains at the forefront of the field.	Research and Development

```

1. .651
location/qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="C230039C09"
/clone_lib="R1EN full-length enriched, 0 day neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory
R1EN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA
primed with a primer [5'
GAGAGAGAGAGAGATCCCAAGAGCTCTTTTTTTTTTTTNN 3'], cDNA
prepared by using trihalose thermo-activated reverse
transcriptase and subsequently enriched for full-length
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 479.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATAATTAATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a Lc
modified pBluescript KS(+) after bulk excision from
FLC 1."
a 207 c 154 g 155 t

```

BASE COUNT	FEC I.	135 a	207 c	154 g	155 t
ORIGIN					

Query Match 34.1%; Score 626.4; DB 9; Length 651;
Best Local Similarity 99.1%; Pred. No. 3.3e-106;
Matches 630; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY	1185	gaagctgaagagcatgagggcacataaagatgctcaccagcgcccccttcagccaggaagga	1244
Db	16	GAAGTTGAAGAGCATGAGGCACATAAAGATGCTACACAGCGCCCTTCAGCAGGGAAGGA	75
QY	1245	ctcgtgcagcctcagcagccagcgctgcctctctctccaccagaattctctctct	1304
Db	76	CTCGTGCAGGTTTCAGCAGCCAGCGGCGCTCTCTTCCTCCACCAAGCATTTTTTCTCTGCT	135
QY	1305	ggctctgtcggatggttaattcgagaactctccaggacaaaactcgggtgtggcacaagg	1364
Db	136	GGTCTTGTGGATGTTAAATTCAGAACTTCCAGGACAAACTCGGGTGTGGCACAAAGG	195
QY	1365	ggctggagccagagccagagccaacgcagagactgcagagagggcaacctgacctaaccc	1424
Db	196	GGCTGGAGCCGACAGACCAGAGCCAGCCAGAGACTGCAGAGAGGCGACCTTCACCTAACCC	255
QY	1425	ccctggaagccaatctgcagttcccggtgtccacccactctctctgagagcgccctcatgc	1484
Db	256	CCCTGGAAGCCAAATCTGCATCTCCGTTGTCCACCCACTCTCTGTAGGAGCGCTCATGC	315

Qy	1485	tctgcccagccctctcccaagggtctacacagagtaaacacacctttttggtccttcgcggtttgggt	1544
Db	316	TCGCCCAGCCCTCTCCCAAGGCTACCAAGATAAACACCTTTTGGCCTTTTCGGTTTGGT	375
Qy	1545	tcttggttcctcaatgaagcctcagagtgctccctcatcgatcttttttgcttttgtcccc	1604
Db	376	TCTTGGTGCTCATATCAGCTCCAGAGTGTCCTCATCGATCTTTTGGCTTTGTGCTCCC	435
Qy	1605	caatccagagggctggaagccatcaccatcattgagagcttaacctgtcagttactagg	1664
Db	436	CAATCCAGGGCTGGAAGCCATCACCATCATTTGGAGCTTAACCTGTCACTTACCTAGG	495
Qy	1665	aggtgctggagagcgccgggttggtttggggtaatacactcactggctctcagccttcta	1724
Db	496	AGGTGCTGGAGGCCCGGGTTGGTTTGGGGTAATCACTCACTGGGCTCTCAGCGCTTCA	555
Qy	1725	acactgagcccttaatacagttccttctgtttggtgactcccaagccccacacaca	1784
Db	556	ACACTGAGCCCTTAATAACAGTTCCTTCTGTGTGGTGACTCCACGCCCCACACACA	615
Qy	1785	cacataaaattatttcgatgctgttttoataactgt	1820
Db	616	CACATAAAATTAATTCGATGCTGTTTCATAACTGT	651

Search completed: May 23, 2002, 23:05:57
Job time: 7513 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:18:10 ; Search time 121.39 Seconds
(without alignments)
3721.229 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacggaggaggaagcgc.....taaaaaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6ptodata/1/ina/PCITUS_COMB.seq:*
6: /cgn2_6ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160.2	8.7	744	4	Sequence 3, Appli
2	160.2	8.7	1512	4	Sequence 1, Appli
3	75.2	4.1	7218	1	Sequence 14, Appl
4	56.2	3.1	7218	1	Sequence 14, Appl
5	55.4	3.0	1931	2	Sequence 2, Appli
6	48.6	2.6	3489	2	Sequence 1, Appli
7	48.6	2.6	32207	2	Sequence 20, Appl
8	48.6	2.6	32207	4	Sequence 20, Appl
9	48.6	2.6	32207	4	Sequence 17, Appl
10	46.8	2.5	289	4	Sequence 17, Appl
11	46.8	2.5	289	4	Sequence 17, Appl
12	44.6	2.4	3319	1	Sequence 2, Appli
13	44.6	2.4	3319	1	Sequence 1, Appli
14	44.6	2.4	3319	2	Sequence 4, Appli
15	44.6	2.4	3319	5	Sequence 2, Appli
16	44.4	2.4	2580	3	Sequence 2, Appli
17	44.4	2.4	2580	4	Sequence 2, Appli
18	44.4	2.4	5452	2	Sequence 1, Appli
19	44.4	2.4	9600	4	Sequence 1, Appli
20	44.4	2.4	10596	1	Sequence 15, Appl
21	44.4	2.4	10596	1	Sequence 15, Appl
22	44.4	2.4	10596	1	Sequence 15, Appl
23	44.4	2.4	10596	1	Sequence 15, Appl
24	44.4	2.4	10596	2	Sequence 15, Appl
25	44.4	2.4	10596	5	Sequence 15, Appl
26	39.2	2.1	24979	2	Sequence 3, Appli
27	39.2	2.1	24979	3	Sequence 3, Appli

c	28	39.2	2.1	24979	5	PCT-US93-03985-3	Sequence 3, Appli
	29	38.2	2.1	1575	3	US-08-478-507-11	Sequence 11, Appl
	30	38.2	2.1	1575	4	US-09-128-275A-11	Sequence 11, Appl
	31	38.2	2.1	16442	3	US-08-781-891-208	Sequence 208, App
c	32	38.2	2.1	53526	3	US-08-658-136-2	Sequence 2, Appli
	33	38.2	2.1	53577	3	US-08-658-136-1	Sequence 1, Appli
c	34	37.6	2.0	16442	3	US-08-781-891-208	Sequence 208, App
	35	37.2	2.0	4112	1	US-08-340-203A-2	Sequence 2, Appli
c	36	37.2	2.0	4112	2	US-08-452-567-2	Sequence 2, Appli
	37	37.2	2.0	4112	3	US-08-452-567-2	Sequence 2, Appli
c	38	37.2	2.0	4112	3	US-09-085-407-2	Sequence 2, Appli
	39	37.2	2.0	4616	1	US-08-340-203A-1	Sequence 1, Appli
c	40	37.2	2.0	4616	2	US-08-452-567-1	Sequence 1, Appli
	41	37.2	2.0	4616	2	US-08-452-567-1	Sequence 1, Appli
c	42	37.2	2.0	4616	3	US-09-085-407-1	Sequence 1, Appli
	43	37.2	2.0	51259	3	US-08-781-891-209	Sequence 209, App
c	44	37	2.0	1433	2	US-08-666-392A-1	Sequence 1, Appli
	45	37	2.0	1433	4	US-09-199-926-1	Sequence 1, Appli

ALIGNMENTS

```

RESULT 1
US-09-163-285-3
; Sequence 3, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: MNI-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)742-4214
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..744
US-09-163-285-3

```

Query Match 8.7%; Score 160.2; DB 4; Length 744;
Best Local Similarity 57.3%; Pred. No. 8.5e-32;

Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
QY 162 gctgagagacgactgggtgcgattgcccagcaaatcgaaagtgcgaagtgtgac 221
DB 75 GGAGGAGGACGATGACAGAGACGCTTGGCCAGCAAAATCGAAGTGTGTAAGCTGCTGAG 134
QY 222 tgtgagctgaagtcgctgtttgagaaacggaagcaagcaaggaagtgattgacacgg 281
DB 135 CACAGAGCTACAGGCGGAAGTCTGAGTCGACCGTGCATCTCGAGAGGTGCTGGAGTGG 194
QY 282 ctatgcatcctggagcgaagggctctgagtcgaag---tacaccaagtcgacttaag 338
DB 195 GCAGGTGCTGGATACAGGCAAGAGGAGACACAGCTGCGCTTACAGGCTTTCAGAGACAAG 254
QY 339 gtttaattgaagtcactgagacatttgcaagaggtcttgactacagcctgacacaaga 398
DB 255 GCTGGAAGAGGCTTAGAGAAATTTATGTAGCGGATCCCTGAGCTATAGTTCACGCTGA 314
QY 399 gaggactggcagcaaccggtttgccaaggttatgtcggagacctttgagacgctgcacaa 458
DB 315 GCGCAAGGCTCACTGAGATATGCCAAGGTCAGAGTCAGACCATGGCAACACTGAAAGG 374
QY 459 cctagtcacaaaggggtcaaggtgctgatatccctctatgagcgtggaagagagac 518
DB 375 CTTAGTGGAGAGTGGTACTTCCACCATCAGGAGCAGCCCTACAAAATTTTCTCTGTGA 554
QY 519 ctacagcagagtgctgacacaaagcagtcgtgctgctggtggaaggtttggaaga 578
DB 435 CACGCTGGAGGTACATACCTCAAGAGGAGTGTGAGACCATGTTGGAGAGTTTGAAGA 494
QY 579 ggtgattgagagtggttacaggaaccaccagggagggaagacctgaactgaattcctctgtgc 638
DB 495 CATTTGGGAGAGTGGTACTTCCACCATCAGGAGCAGCCCTACAAAATTTTCTCTGTGA 554
QY 639 caaccagtgctgaaggaagacagagtgctcctagcagacgctgctggaagaa 698
DB 555 AGGTCATGTGCTCCAGCTGCTGAAACTGTCATGCTACAGGAAACTTGGACTGGAAAGGA 614
QY 699 g 699
DB 615 G 615

RESULT 2
US-09-163-285-1
; Sequence 1, Application US/09163285
; Patent No. 6204013
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MSP-5 PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,285
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/090,398
; FILING DATE: June 24, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: MNI-049
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1512 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 125..868
US-09-163-285-1

Query Match 8.7%; Score 160.2; DB 4; Length 1512;
Best Local Similarity 57.3%; Pred. No. 1.1e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
QY 162 gctgagagacgactgggtgcgattgcccagcaaatcgaaagtgcgaagtgtgac 221
DB 199 GGAGGAGGACGATGACAGAGACGCTTGGCCAGCAAAATCGAAGTGTGTAAGCTGCTGAG 258
QY 222 tgtgagctgaagtcgctgtttgagaaacggaagcaaggaagtgattgacacgg 281
DB 259 CACAGAGCTACAGGCGGAAGTGTGAGTCGACCGTGCATCTCGAGAGGTGCTGGAGTGG 318
QY 282 ctatgcatcctggagcgaagggctctgagtcgaag---tacaccaagtcgacttaag 338
DB 319 GCAGGTGCTGGATACAGGCAAGAGGAGACACACGTGCTTACAGCGCTTTCAGAGACAAG 378
QY 339 gtttaattgaagtcactgagacacatttgcaagaggtcttggagacacctgacacagga 398
DB 379 GCTGGAAGAGGCTTAGAGAAATTTATGTAGCGGATCCCTGAGCTATAGTGTTCACGCTGA 438
QY 399 gaggactggcagcaaccggtttgccaaggttatgtcggagacaccttggagacgctgcacaa 458
DB 439 GCGCAAGGCTCACTGAGATATGCCAAGGTCAGAGTCAGACCATGGCAACACTGAAAGG 498
QY 459 cctagtcacaaaggggtcaaggtgctgatatccctctatgagcgtggaagagagac 518
DB 499 CTTAGTGGAGAGGCTTAGAGTGTGAGTCTGGGATCCCTCTGAGAGCTTGGGATGAGCC 558
QY 519 ctacagcagagtggtgacctcaaggaagcagtcgtgctggtggaagaggtttggaaga 578
DB 559 CACGCTGGAGGTACATACCTCAAGAGGAGTGTGAGACCATGTTGGAGAGTTTGAAGA 618
QY 579 ggtgattgagagtggttacaggaaccaccagggagggaagacctgaactgaattcctctgtgc 638
DB 619 CATTTGGGAGAGTGGTACTTCCACCATCAGGAGCAGCCCTACAAAATTTTCTCTGTGA 678
QY 639 caaccagtgctgaaggaaggaagcagagtgctcctagcagagcgtggtgctggaagaa 698
DB 679 AGGTCATGTGCTCCAGCTGCTGAAACTGTCATGCTACAGGAAACTTGGACTGGAAAGGA 738
QY 699 g 699
DB 739 G 739

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52

```
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEX:      (703)683-4109
;   TELETYPE:   899149
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH:       7218 base pairs
; TYPE:         nucleic acid
; STRANDEDNESS: single
; TOPOLOGY:     linear
; IMMEDIATE SOURCE:
; CLONE:        pTZgpt-FIs
US-08-232-463-14

Query Match          3.1%; Score 56.2; DB 1; Length 7218;
Best Local Similarity 7.2%; pred. No. 7.4e-05;
Matches    31; Conservative 223; Mismatches 180; Indels    0; Gaps    0;

Qy  1405 gagggacccctgcactaaccccccttgaaagccaatcctgcagtcccgctgtcccaccacctc 1464
      ||||| | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : ;
Db  1052 GAGGGACGCTTGCATYYYYYYYYVYYYYYYVYYYYVVVVVVVVVVVVVVVVVVVVVVVVV 1111

Qy  1465 ctctcaggagcgcctcatgtctgccagcccttctcccagggcctaccagagtaaacacc 1534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : ;
Db  1112 YYYYYYYYVYYYYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1171

Qy  1525 ttctggccttcggtttggcttcctgcctcatcacgccctccagagtgctccctcatcga 1584
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : ;
Db  1172 YYYYYYYYVYYYYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1231

Qy  1585 tcttttgtcctttgtcccceaatcccaaggcggctggaagccatcaccatcatattggaggc 1644
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : ;
Db  1232 YYYYYYYYVYYYYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1291
```

```

Qy 1645 ttaacgtctcagttactagaggtgctggagcgccgggtgtgttgggtggtatcact 1704
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1292 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1351

Qy 1705 cactggctctcagcctcttaacactgagcccttaataacagttcctctgttggtga 1764
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1352 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1411

Qy 1765 ctccacgccccacacacacacataaattattcgtgctgtttcataactgtaaaa 1824
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1412 YYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAATCTCTATCTCTTAATACTTGCATAG 1471

Qy 1825 aaaaaaaaaaaaaa 1837
   || || ||
Db 1472 ATAGGTAATACA 1484

```

```

RESULT 5
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

```

Query Match 3.0%; Score 55.4; DB 2; Length 1931;
Best Local Similarity 46.5%; Pred. No. 7.2e-05;
Matches 179; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 463 gtccacaagggtcaaggtgtgatgatatccctctatgactgtggaagacacctca 522
   || || || || || || || || || || || || || || || || || || || || ||
Db 940 GACGAGGACGGGAGGACGGGAGGACGAGGACGGGAGGACGGGAGGACGGGAGGACGGG 881

Qy 523 gcagaggtggtgacctcaagaagcagtgtagctgtggtggaagatttgaagaggtg 582
   || || || || || || || || || || || || || || || || || || || || ||
Db 880 GAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGGG 821

Qy 583 attgaggtggttacaggaacaccagaggaagacacctgactgaattctctgtgccaac 642
   || || || || || || || || || || || || || || || || || || || || ||
Db 820 GAGGAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGAG 761

Qy 643 cactgtctgaaggaaagacacagtgcttcctagcagagcggtgtgtggaagagggg 702
   || || || || || || || || || || || || || || || || || || || || ||
Db 760 GACGGGAGGACGGGAGGACGAGGACGGGAGGACGAGGACGGGAGGACGAGGACGGGAGG 701

Qy 703 gacatagctccctggggaggaagaaatccaaagaagacgagcagtgagtcgaaggctcc 762
   || || || || || || || || || || || || || || || || || || || || ||
Db 700 GAGGACGAGGACGGGAGGACGGGAGGACGGGAGGAGGAGGAGGACGAGGACGGGAGGACGAG 641

Qy 763 tccagtgcagcagcagcagaggaagaaactgggggacctgggggagatgccaagccc 822
   || || || || || || || || || || || || || || || || || || || || ||
Db 640 GACGGGAGGACGGGAGGACGGGAGGACGGGAGGACGGGAGGAGGAGGACGAGGACGGGAG 581

Qy 823 gaggaggaggagggtgtgcagaag 847
   || || || || || || || || || || || || || || || || || || || || ||
Db 580 GACGGGAGGACGGGAGGACG 556

```

```

RESULT 6
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

```

```

Query Match 2.6%; Score 48.6; DB 2; Length 3489;
Best Local Similarity 48.4%; Pred. No. 0.0049;
Matches 194; Conservative 0; Mismatches 204; Indels 3; Gaps 2;

Qy 446 agacgtgcacacacctagctccacaaaggggtcaaggtggtgtagtatccctatgac 505
   || || || || || || || || || || || || || || || || || || || || ||
Db 2243 AGCAGCAGGATGAACAGGAGCAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGG 2302

Qy 506 tgtggaacgagacctcagcagaggtggtgacctcaagaagcagtgtagctgtgctgtgg 565
   || || || || || || || || || || || || || || || || || || || || ||
Db 2303 AGTTAGAGCAGCAGGAGCAGGAGGAGTTAGAGGATCAGGACAGGAGTTAGAGCAGGAGCAG 2362

Qy 566 aagagttgaagaggtgattgagagctggttacaggaacccacagaggagaaacctgactg 625
   || || || || || || || || || || || || || || || || || || || || ||
Db 2363 AGGAGTTAGAGGAGCAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAGGAGTTAGAGGAGCAGG 2422

Qy 626 aattctctgtgccaaccacgtgctgaagggaagaaagacacagagttgctcagacagcgt 685
   || || || || || || || || || || || || || || || || || || || || ||
Db 2423 AGCAGGAGTTAGAGGAGCAGGAGGAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGGAG 2480

Qy 686 ggtctggaagaagggggacatagcctcctctggaggggaagaaatcccaagaagaagcgca 745
   || || || || || || || || || || || || || || || || || || || || ||
Db 2481 CGAGGAGCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGGAGCA 2540

```



```

Qy 566 aagagttgaagaggtgattgactggttacaggaaacccaccagggaagaaacactgactg 625
    ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19634 AGGAGTTAGAGGACGAGGACGAGGAGTTAGAGGACGAGGAGGAGTTAGAGGACGAGG 19575
Qy 626 aattcctctgtgccaaccacgtctgaagggaaggaaggaaggaaggaaggaaggaag 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19574 AGCAGGAGTTAGAGGACGAGGACGAGGAGTTAGAGGAGC--AGGAGCAGGAGTTAGAGGA 19517
Qy 686 ggtctgccaagaaggggacatagcctccctggggagggaagaaatccaaagaagacgca 745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19516 GCAGGAGCAGGAGTTAGAGGACGAGGACGAGGAGTTAGAGGAGCAGGAGTTAGAGGAGCA 19457
Qy 746 gcgagtcaggggtccctccctgagtcagcagcagcagcagcagcagcagcagcagcag 805
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19456 AGACAGGAGTTGGAGAGCAAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19398
Qy 806 gggaggtgccaagcccgagagggagggggtgtgcagaag 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19397 AGGACAGAGCAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19357

RESULT 9
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 2.6%; Score 48.6; DB 4; Length 32207;
Best Local Similarity 48.4%; Pred. No. 0.012;
Matches 194; Conservative 0; Mismatches 204; Indels 3; Gaps 2;

Qy 446 agacgtcacacacactagtcacaaagggtcaaggtgtgatgatatccctatgac 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19754 AGCAGCAGATGATAGAGGACGACGAGGAGGACGAGGACGAGGAGGAGGAGGAGGAGG 19695
Qy 506 tdtggaacgagacacctcagcagaggtggtgacctcaagaagcagtgatgactgctgtg 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19694 AGTTAGAGGACGAGGACGAGGAGTTAGAGGATCAGGACGAGGAGTTAGAGGAGCAGGAGC 19635
Qy 566 aagagttgaagaggtgattgaggactggtacaggaaacccaccagggaagaaacactgactg 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19634 AGGAGTTAGAGGACGAGGACGAGGAGTTAGAGGACGAGGAGGAGGAGGAGTTAGAGGAGCAGG 19575
Qy 626 aattcctctgtgccaaccacgtctgaagggaaggaaggaaggaaggaaggaaggaagga 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19574 ACAGCAGGTTAGAGGACGAGGAGGAGGAGTTAGAGGAGC--AGGAGCAGGAGTTAGAGGA 19517
Qy 686 ggtctgccaagaaggggacatagcctccctggggagggaagaaatccaaagaagacgca 745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19516 GCAGGAGCAGGAGTTAGAGGACGAGGAGGAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCA 19457
Qy 746 gcgagtcaggggtccctccctgagtcagcagcagcagcagcagcagcagcagcagcag 805
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 19456 AGACAGGAGGTGGAAGCAAGACGAGCAGGACGAGGAGGAGGAGGAGGAGGAGGAGG 19398
Qy 806 gggaggtgccaagcccgagagggagggggtgtgcagaag 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19397 AGGACAGACAGCAGGACGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19357

RESULT 10
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 2.5%; Score 46.8; DB 4; Length 289;
Best Local Similarity 5.3%; Pred. No. 0.0054;
Matches 11; Conservative 101; Mismatches 96; Indels 0; Gaps 0;

Qy 640 aaccacgtgctgaaggaaagacacagagtccttagcagagcgtgtgctggaagaag 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 rarcrararurgnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 104
Qy 700 ggggacatagcctccctggagggaagaaatccaaagaagacgagcagcagcagcagcag 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 rnrrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 164
Qy 760 tectcagtgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 819
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165 rnrrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 224
Qy 820 gccgaggaagaggggtgtgcagaag 847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 225 rcrurgrcrurarcrcrururur 252

RESULT 11
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796

```

TELEFAX: (206) 236 0205
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Leishmania chagasi
US-08-006-676B-2

Query Match	2.4%;	Score 44.6;	DB 1;	Length 3319;
Best Local Similarity	46.1%;	Pred. No. 0.051;		
Matches 149;	Conservative 0;	Mismatches 174;	Indels 0;	Gaps 0;
QY	419	ttgcgaagggtatgtcggagagacctttgagcgcgtgcacaacctagtccacaagggtca	478	
Db	1629	TTGTGATGTTTGGCAGGTGACAGAGACCCGCGGCACCGGGATCCGGGAGCTGGAGG	1688	
QY	479	aggtgggtgatggatatccctctatgagctgtggaaacgagacctcagcagaggttgctgacc	538	
Db	1689	AGCAGATGGAGGACATCGCGCAGCGATGCTGGCGCGCACCCGCGTACGTGCTCGAC	1748	
QY	539	tcaagaagcagttgacgtgtcctggtggaagagtttgaaagagtgattgaggaactggtaca	598	
Db	1749	TGAAGAGAAGAGCTTTCGCTCTGGAGTCGGAGGCGCAGACGCTGCGCGGAGCTGCAGG	1808	
QY	599	ggaaccaccagggaagaaacctgactgaattctctctgtgccaaaccacgtgctgaaggaa	658	
Db	1809	CGCTGGAGAGCGGACGGGGAGCAACACCGTGGCAGGACGGCTGCTCGCGCGCAGCGGAGG	1868	
QY	659	aggacacaggttgcctagcagacggttggtcttgcaagaagggggacatagctccctgg	718	
Db	1869	CGGAGAAGACCGAGCTGGAGTCGGCTGCGGTGCGCTGCGAGGAGGATGACCCGGGACTC	1928	
QY	719	gagggaagaataccagaagaag	741	
Db	1929	GACGGCAGCGGACACAGATGCAG	1951	

```

RESULT 13
US-08-282-845-1
; Sequence 1, Application US/08282845
; Patent No. 5719263
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1
; SOFTWARE: Microsoft Word for Macintosh 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/006,676
; FILING DATE: JANUARY 15, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne

```

; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: 5004-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: rk39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 455..3319
; US-08-282-845-1

Query Match 2.4%; Score 44.6; DB 1; Length 3319;
Best Local Similarity 46.1%; Pred. No. 0.051;
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 419 ttgccaagggtatgtcggagacacctttgagacgctgcacaaacctagtcacaaagggtca 478
DB 1629 TTGTGAATGTTCCGAGGTGAACGAGGACCCGCGGCACACGGCGGATCCGCGAGCTGGAGG 1688

QY 479 agtgtgtgatgatccctctagctgtggaagagcttcagagagacctcagagaggttgctgacc 538
DB 1689 ACACATGAGGACATGCGCGCAGCGCATGGCTGGCGCGCACCCCGGTACGTGCTTGAGC 1748

QY 539 tcaagaagcagtgtagctgtcgtgtggaagagtttgaagaggtgattgaggactggtaca 598
DB 1749 TGAAGAAGAGCTTGGCTGCTGGAGTCGGAGCGGACAGCGGCGGACCTGCAGG 1808

QY 599 ggaaccaccagagagaagacctgaattcctctgtgccaaaccagtgctgaaggga 658
DB 1809 CGCTGGAGAGGAGCGGGAGCACACACAGGTGCAGGCGGCTGCTGCGCGCGAGG 1868

QY 659 agcacagagtgctagcagagcgtgtgtggaagaggggacatagcctccctgg 718
DB 1869 CGGAGAAGAGCGAGCTGGAGTCGGCTGCGGCTGCGCTGCGAGGAGGATGACCGGACTC 1928

QY 719 gagggagaagaatcccaagaagaag 741
DB 1929 GACGGCAGCGGACAAAGATGCAG 1951

RESULT 14
PCT-US94-414A-4
; Sequence 4, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: LEISHMANIASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/428,414A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Kadlecak, Ann T.
; REGISTRATION NUMBER: 39,244
; REFERENCE/DOCKET NUMBER: 210121.407
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-428-414A-4

Query Match 2.4%; Score 44.6; DB 2; Length 3319;
Best Local Similarity 46.1%; Pred. No. 0.051;
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 419 ttgccaagggtatgtcggagacacctttgagacgctgcacaaacctagtcacaaagggtca 478
DB 1629 TTGTGAATGTTCCGAGGTGAACGAGGACCCGCGGCACACGGCGGATCCGCGAGCTGGAGG 1688

QY 479 agtgtgtgatgatccctctagctgtggaagagcttcagagagacctcagagaggttgctgacc 538
DB 1689 ACACATGAGGACATGCGCGCAGCGCATGGCTGGCGCGCACCCCGGTACGTGCTTGAGC 1748

QY 539 tcaagaagcagtgtagctgtcgtgtggaagagtttgaagaggtgattgaggactggtaca 598
DB 1749 TGAAGAAGAGCTTGGCTGCTGGAGTCGGAGCGGCGGACAGCGGCGGACCTGCAGG 1808

QY 599 ggaaccaccagagagaagacctgaattcctctgtgccaaaccagtgctgaaggga 658
DB 1809 CGCTGGAGAGGAGCGGGAGCACACACAGGTGCAGGCGGCTGCTGCGCGCGAGG 1868

QY 659 agcacagagtgctagcagagcgtgtgtggaagaggggacatagcctccctgg 718
DB 1869 CGGAGAAGAGCGAGCTGGAGTCGGCTGCGGCTGCGCTGCGAGGAGGATGACCGGACTC 1928

QY 719 gagggagaagaatcccaagaagaag 741
DB 1929 GACGGCAGCGGACAAAGATGCAG 1951

RESULT 15
PCT-US94-00324-2
; Sequence 2, Application PC/TUS9400324
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven
; TITLE OF INVENTION: Diagnosis of Leishmaniasis
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple System 7.1
; SOFTWARE: Microsoft Word, version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00324
; FILING DATE:
; CLASSIFICATION:

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/006,676
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: 5004-WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Leishmania chagasi
; PCT-US94-00324-2

Query Match      2.4%; Score 44.6; DB 5; Length 3319;
Best Local Similarity 46.1%; Pred. No. 0.051;
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 419 ttgccaagggtatgtcgagacacatttgagagcgtgcacaaacctagtcacacaaaggggtca 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1629 TTGTGAATGTTGGCGCAGGTGAACGAGGACCGCCGCCGACGGCGGATCCGGGAGCTGGAGG 1688

QY 479 aggtggtgatggatatccctctatgagctgtggaacgagacctcagcagaggtggtgacc 538
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1689 AGCAGATGGAGGACATCGCGCAGCGGATGGCTGGCGCGGACCCCGCTACGTGTCTGAGC 1748

QY 539 tcaagaagcagtgtagcgtgtgtgtggaagagtttgaagaggtgattgaggactggtaca 598
   . ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1749 TGAAGAAGAAAGCTTGCCTGTGCGAGTCGGAGGCGCAGAAAGCGTGGCGCGGACCTGCAGG 1808

QY 599 ggaaccaccaggaagagacctaactgaattcctctgtgccaaaccagtgctgaaggaa 658
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1809 CGTGGAGAGGGGCGGCGGACACACAGGTGCAGGAGCGGCTGCTGGCGCGGACGAGG 1868

QY 659 agaacacagttgctctagcagagcgggtggtcttgcaagaagggggacatagcctccctgg 718
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1869 CGGAGAAGAGCGAGCTGGAGTCGGCTGCGGCTGCGGCTGCAGGAGGAGATGACCGCGACTC 1928

QY 719 gagggaagaataccaagaagaag 741
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1929 GACGGCAGCGCGGACAGATGCAG 1951
```

Search completed: May 24, 2002, 00:20:16
Job time: 11682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:23:11 ; Search time 525.6 Seconds
(without alignments)
6007.241 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgagggaggaagcgcc.....taaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues 3472872
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	1839	100.0	1839	22	AAS05278	Murine trinucleotide
2	1725.2	93.8	1848	22	AAS05286	DNA sequence of an
3	691.6	37.6	1675	22	AAS21311	Human cDNA sequenc
4	665.8	36.2	1362	22	AAS05279	Human trinucleotide
5	446	24.3	1657	20	AA04377	Human secreted pro
6	412.2	22.4	580	22	AAS25051	Human ovarian PCR-
7	412.2	22.4	580	22	AAH83696	Human ovarian tumo
8	224.4	12.2	274	21	AAA42779	Human secreted exp
9	160.2	8.7	744	21	AA249678	Human myocardium s

10	160.2	8.7	746	22	AAF94104	Primer specific fo
11	160.2	8.7	747	22	AAC90704	Human secretory pr
12	160.2	8.7	1419	22	AAF93914	Human cDNA encodin
13	160.2	8.7	1457	20	AAF78923	Human zcalc-1 DNA
14	160.2	8.7	1462	21	AA16621	Human secreted pro
15	160.2	8.7	1470	21	AA196349	CDNA encoding a no
16	160.2	8.7	1470	21	AAC58116	Human PRO4354 nucl
17	160.2	8.7	1470	21	AAC58116	Human DNA encoding
18	160.2	8.7	1512	21	AAZ49677	Human myocardiolum s
19	156.6	8.5	1658	21	AAZ98029	Human secreted pro
20	156.6	8.5	1658	21	AAZ98029	Human secreted pro
21	151.4	8.2	750	23	ABL16485	Drosophilla melanog
22	151.4	8.2	750	23	ABL16485	Drosophilla melanog
23	151.4	8.2	750	23	ABL16485	Drosophilla melanog
24	142.4	7.7	189	16	AAT23979	Human gene signatu
25	137	7.4	1623	20	AA88064	Human zcalc-1 DNA
26	123.6	6.7	1306	20	AA224902	Human secreted pro
27	123.6	6.7	1312	22	AA05462	Human secreted pro
28	117.8	6.4	1315	22	AA05464	Human secreted pro
29	107	5.8	107	22	AAS05289	DNA sequence of de
30	98.6	5.4	293	20	AA232038	Human METH2 relate
31	98.6	5.4	293	22	AAC90095	HTOFC34RA cDNA clo
32	69.4	3.8	4069	23	ABL11794	Drosophilla melanog
33	67.6	3.7	501	22	AAH29134	Drosophilla melanog
34	67.6	3.7	510	22	AAH29134	Drosophilla melanog
35	66	3.6	471	22	AAS05284	Umbilical vein end
36	57.4	3.1	10732	21	AAAL0594	T243 gene homolog
37	51	2.8	443	20	AAH8063	Gene encoding a su
38	49.4	2.7	1416	23	AAS71051	Human zcalc-1 DNA
39	49.4	2.7	4720	24	AA594782	DNA encoding novel
40	49.4	2.7	5240	22	AA52624	Human DNA sequence
41	49.4	2.7	5360	22	AAK51640	Human polynucleoti
42	48.6	2.6	3489	21	AAK30290	Human polynucleoti
43	48.6	2.6	3489	22	AAH82901	Kaposi's sarcoma-a
44	48.6	2.6	32207	20	AAH73805	Nucleotide sequenc
45	48.6	2.6	137507	19	AAV19941	KSHV LTR DNA (nucl
						KSHV long unique c

ALIGNMENTS

RESULT 1
AAS05278
ID AAS05278 standard; cDNA; 1839 BP.
AC AAS05278;
XX
XX
DT 07-SEP-2001 (first entry)
DE Murine trinucleotide repeat protein (TRP) cDNA sequence.
DE Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; murine; ss.
XX Mus sp.
XX
XX Key Location/Qualifiers
FT CDS 55..885
FT FT /*tag= a
FT FT /product= "TRP"
XX
XX WO200130798-A1.
XX
XX PD 03-MAY-2001.
XX
XX PF 26-OCT-2000; 2000WO-US29382.
XX
XX PR 26-OCT-1999; 99US-0161488.
XX (DELT-) DELTAGEN INC.
XX
XX PI Klein R, Matthews W, Moore M, Allen KD;

QY	892	cttagtgctcttgaa	tcaagacccctgact	tcagagcttgaggacacacagcagcgc	951
Db	901	cttagtgctcttgaa	tcaagacccctgact	tcagagcttgaggacacgacagcagcgc	960
QY	952	agcgcagctccacga	gcagcagctgct	tcagacatcaggtctcctcccttgactgccc	1011
Db	961	agcgcagctccacga	gcagcagctgct	tcacagatcaggctcctcccttgctgctgccc	1020
QY	1012	cctttctcctcttgaa	cacagcagaggttggaagatctggggtgctggagacggca	1071	
Db	1021	cctttctcctcttgaa	cacagcagaggttggaagatctggggtgctggagacggca	1080	
QY	1072	ccccaaaggaagag	gagagcagagcagctctctttctacacagtcctccctcac	1131	
Db	1081	ccccaaaggaagag	gagagcagagcagctctctttcacacagtcctccctcac	1140	
QY	1132	gagctccggggttcc	accacagctccccaggtcgagatccaggtccttgacatggaagctg	1191	
Db	1141	gagctccggggttcc	accacagctccccaggtcgagatccaggtccttgacatggaagctg	1200	
QY	1192	aagacatgaggcaca	taagatgctcacagcgcccccttcagccaggaaggactccgtg	1251	
Db	1201	aagacatgaggcaca	taagatgctcacagcgcccccttcagccaggaaggactccgtg	1260	
QY	1252	cagctccagcagcag	gcctgccttctctccacaaagattctcttgcttggtcctt	1311	
Db	1261	cagctccagcagcag	gcctgccttctctccacaaagattctcttgcttggtcctt	1320	
QY	1312	gtcggatgtaaatct	gagaacttcaggacaaactcgggtgtggcacaagggctgga	1371	
Db	1321	gtcggatgtaaatct	gagaacttcaggacaaactcgggtgtggcacaagggctgga	1380	
QY	1372	cgccagacgacgac	cgccagagactgcagagaggcgaactgaactaacccccctgga	1431	
Db	1381	cgccagacgacgac	cgccagagactgcagagaggcgaactgaactaacccccctgga	1440	
QY	1432	aagcgaactcgag	ttcccggtgccaccgaactcctcttgaggacgcctcatgctctgccc	1491	
Db	1441	aagcgaactcgag	ttcccggtgccaccgaactcctcttgaggacgcctcatgctctgccc	1500	
QY	1492	agccctctccaggg	ctaccagagtaaacacttttgccctttcggtttggttcctggg	1551	
Db	1501	agccctctccaggg	ctaccagagtaaacacttttgccctttcggtttggttcctggg	1560	
QY	1552	tctctatcagctcc	agagtgctcccctcatcgacttttttgccctttgctccccaatccc	1611	
Db	1561	tctctatcagctcc	agagtgctcccctcatcgacttttttgccctttgctccccaatccc	1620	
QY	1612	aggggctggaagg	ccatcaccaatcattggaggcttaacctgctagttactagagtgct	1671	
Db	1621	aggggctggaagg	ccatcaccaatcattggaggcttaacctgctagttactagagtgct	1680	
QY	1672	ggagcgcccggggt	gtgttggtgataactcactcactgctctcagcctctctaaactgc	1731	
Db	1681	ggagcgcccggggt	gtgttggtgataactcactcactgctctcagcctctctaaactgc	1740	
QY	1732	agcccttaatacag	tctctctgtgtgtgctcccaagccccccacacacacata	1791	
Db	1741	agcccttaatacag	tctctctgtgtgtgctcccaagccccccacacacacata	1800	
QY	1792	aaattatttcgat	gtgtttcataactgtaaaaaataaaaaa	1839	
Db	1801	aaattatttcgat	gtgtttcataactgtaaaaaataaaaaa	1848	

XX	Human cDNA sequence encoding for PRO4409 polypeptide.
DE	
XX	
KW	Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW	breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW	cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW	adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
XX	WO200140466-A2.
PN	
XX	
PD	07-JUN-2001.
XX	
PF	01-DEC-2000; 2000WO-US32678.
XX	
PR	01-DEC-1999; 99WO-US28301.
PR	01-DEC-1999; 99WO-US28634.
PR	02-DEC-1999; 99WO-US28551.
PR	02-DEC-1999; 99WO-US28564.
PR	02-DEC-1999; 99WO-US28565.
PR	09-DEC-1999; 99US-0170262.
PR	16-DEC-1999; 99WO-US30095.
PR	20-DEC-1999; 99WO-US30911.
PR	20-DEC-1999; 99WO-US30999.
PR	30-DEC-1999; 99WO-US31243.
PR	06-JAN-2000; 2000WO-US00277.
PR	06-JAN-2000; 2000WO-US00376.
PR	11-FEB-2000; 2000WO-US03565.
PR	18-FEB-2000; 2000WO-US04341.
PR	18-FEB-2000; 2000WO-US04342.
PR	22-FEB-2000; 2000WO-US04414.
PR	24-FEB-2000; 2000WO-US04914.
PR	24-FEB-2000; 2000WO-US05004.
PR	01-MAR-2000; 2000WO-US05601.
PR	20-MAR-2000; 2000WO-US07377.
PR	21-MAR-2000; 2000WO-US07532.
PR	30-MAR-2000; 2000WO-US08439.
PR	17-MAY-2000; 2000WO-US13705.
PR	22-MAY-2000; 2000WO-US14042.
PR	30-MAY-2000; 2000WO-US14941.
PR	02-JUN-2000; 2000WO-US15264.
PR	10-NOV-2000; 2000WO-US30873.
XX	
PA	(GETH) GENENTECH INC.
XX	
XX	Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI	Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI	Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX	
XX	WPI; 2001-408281/43.
DR	P-PSDB: AAU1239.
DR	

RESULT	3
AAS21311	
ID	AAS21311 standard; cDNA; 1675 BP.
XX	
AAS21311;	
XX	
DT	24-OCT-2001 (first entry)

CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or adipocytes; or inhibit binding of A-peptide
CC to factor VIIa. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 other:

Query Match	37.6%	Score 691.6;	DB 22;	Length 1675;
Best Local Similarity	84.5%;	Pred. NO. 4.5e-169;		
Matches 791; Conservative	0;	Mismatches 139;	Indels 6;	Gaps 1;

Qy	5	c	g	a	g	a	g	a	g	a	g	c	c	c	g	g	t	c	g	c	t	g	c	t	g	c	t	g	g	t	c	g	c	t	g	g	t	c	c	a	t	g	a	g	c	c	a		64
Db	287	c	g	g	a	g	a	g	a	a	c	c	c	c	g	t	c	c	t	t	a	g	g	t	c	c	g	c	c	g	c	c	g	c	c	g	c	c	a	t	g	a	t	c	a		346		
Qy	65	t	g	t	c	a	g	c	t	c	g	c	c	c	g	t	g	c	t	t	a	t	t	c	t	t	t	g	t	g	c	t	g	c	t	g	c	t	g	c	t	g	c	t	g		124		
Db	347	t	g	c	t	c	g	a	c	c	g	t	c	c	g	t	g	c	t	g	t	c	t	t	c	c	t	t	g	c	t	g	c	t	g	c	t	g	c	t	g	c	t	g		406			
Qy	125	t	c	c	t	c	t	g	c	c	c	a	a	g	t	a	g	c	c	c	g	a	g	t	c	c	c	g	g	c	g	t	a	g	a	g	a	c	c	a	c	t	g	g		184			
Db	407	t	g	c	t	c	g	c	g	c	c	c	a	g	t	g	c	c	c	a	g	c	c	c	c	c	c	c	c	a	g	c	c	c	a	g	a	c	a	c	c	t	g		466				
Qy	185	g	a	t	t	g	c	c	a	a	a	t	g	c	a	a	g	t	g	c	a	a	t	a	t	g	t	g	c	t	g	a	g	c	t	g	a	a	c	t	g	g		244					
Db	467	g	c	t	g	c	c	a	a	a	t	g	c	a	a	t	g	t	g	c	a	a	t	a	t	g	t	g	c	t	g	a	a	c	t	g	a	a	c	t	g		526						
Qy	245	a	g	a	a	c	g	a	a	a	g	a	a	g	t	a	t	g	a	c	a	c	c	g	c	t	a	t	a	g	c	a	c	c	c	t	a	c	c	t	g	a	c		304				
Db	527	a	g	a	a	c	g	c	a	a	c	c	a	a	g	a	g	t	g	a	t	g	c	a	c	c	c	c	c	c	a	a	c	c	c	a	a	c	c	c	a	a		586					
Qy	305	g	c	t	g	a	t	c	a	a	g	t	a	c	a	a	g	t	a	c	c	g	t	a	a	t	a	a	g	t	a	a	c	t	g	a	a	c	c	a	t		364						
Db	587	c	c	t	g	a	t	c	a	a	t	a	c	a	a	g	t	c	g	a	c	t	c	g	g	t	a	a	t	c	a	a	c	c	c	a	a	c	c	a	t		646						
Qy	365	g	a	a	a	g	c	t	t	g	a	c	t	a	c	a	a	g	a	g	a	g	a	c	t	g	c	a	a	a	g	a	c	c	c	a	a	c	c	c	t	g	c	a		424			
Db	647	g	c	a	a	a	g	c	t	g	a	t	a	g	c	t	g	c	a	a	a	g	a	g	a	c	c	g	a	c	c	a	a	c	a	a	c	a	a	c	a		706						
Qy	425	a	g	g	t	a	t	c	g	a	c	t	t	g	a	c	t	g	c	a	a	c	t	a	c	a	a	c	t	a	c	a	a	a	g	g	c	c	a	a	g		484						
Db	707	a	g	g	c	a	t	g	c	a	a	c	t	t	g	a	c	a	c	c	c	g	t	a	c	a	a	c	c	c	g	t	a	c	a	a	a	g	g	c	c	a		766					
Qy	485	t	g	a	t	a	t	c	c	c	t	a	g	a	c	t	g	t																															

Db	1127	tcagaaggcatccctctctcacacagacccccctgatgctctgagccccccagcat	1186
Qy	899	tccttgatcaagaccctgacttcagacttgga	934
Db	1187	ccctgtcctgagaccctgattttgaagctgagga	1222
RESULT	4		
AA505279			
ID	AA505279	standard; cdna; 1362 BP.	
XX			
AC	AA505279;		
XX			
DT	07-SEP-2001	(first entry)	
XX			
DE	Human trinucleotide repeat protein (TRP)	cdna sequence.	
XX			
KW	Trinucleotide repeat protein; TRP; T243;	embryonic stem cell; ES;	
KW	transgenic animal; knockout mouse;	triplet repeat expansion;	
XX	fragile X syndrome; Huntington's disease;	human; ss.	
OS	Homo sapiens.		

RESULT 4
AAS05279
ID AAS05279 standard; cDNA; 1362 BP.

XX	
AC	AAS05279;
XX	
DT	07-SEP-2001 (first entry)

Human trinucleotide repeat protein (TRP) cDNA sequence.

KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; human; ss.

OS Homo sapiens.

key	Location/Qualifiers
CDS	7..843
FT	/*tag= a
FT	/product= "TRP"

PN WO200130798-A1.

03-MAY-2001.

26-OCT-2000: 2000WO-US29382

AA 26-OCT-1999; 99US-0161488.

AA (DELT-) DELTAGEN INC.

PI Klein R, Matthews W, Moore M, Allen KD:

WPI; 2001-300473/31.

XX
F-F3DB, AA002433.

PT novel transgenic function of a gene

XX
XX

XX
XX
XX

(TRP). The invention describes methods of producing embryonic stem (ES) cells comprising a heterozygous disruption in a target DNA sequence (preferably gene T423) encoding a TRP and of producing a knockout mouse comprising a homozygous disruption in a gene encoding TRP. The invention also relates to identifying agents capable of affecting a phenotype of a knockout mouse. Also described are methods of determining whether expansion of the trinucleotide repeat in a gene encoding TRP produces a phenotypic change. The transgenic animals and the cells are useful for identifying compounds capable of ameliorating disease symptoms, and as test substrates for the identification of drugs, pharmaceuticals, therapies and interventions which may be effective in treating trinucleotide repeat disorders e.g. fragile x syndrome and Huntington's disease. The animal models for trinucleotide repeat disorders are ideal model systems to study the progression of disease *in vivo*, the molecular basis of these diseases and show the features observed in human disease using the mice, it is possible to model both the pathogenic mechanism and the trinucleotide repeat instability in the mouse.

Sequence 1362 BP; 352 A; 376 C; 392 G; 242 T; 0 other;

Query Match

36.28; Score 665.8; DB 22; Length 1362;

[illegible]


```
Db 361 gtacaggaaccaccaggaggagaaacctgactgaattcctctgcgcacaccacgtgtgtgaa 420
Qy 654 ggaagaagacacagattgctctacagacggtggtctgcaagaaggaggacacatagcctc 713
Db 421 gggaaaadacacacagttgctgcgcagagcagtggctccggcaagaaggaggagacacagacct 480
Qy 714 cctgggagg 723
Db 481 gcccgggcgg 490

RESULT 7
AAH83696
ID AAH83696 standard; cDNA; 580 BP.
XX
AC AAH83696;
XX
DT 25-SEP-2001 (first entry)
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1320.
XX
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
KW immunogenic; vaccine; ss.
OS Homo sapiens.
XX
PN WO200151513-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01575.
XX
PR 14-JAN-2000; 2000US-0176722.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA;
XX
DR WPI; 2001-425866/45.
XX
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to
PT treat and diagnose cancers, particularly ovarian cancer -
XX
PS Claim 5; Page 304; 338pp; English.
XX
CC AAH82377 to AAH83878 represent human ovarian tumour-associated
CC polynucleotide sequences which encode ovarian tumour proteins. The
CC ovarian tumour protein and polynucleotide sequences have cytostatic
CC activity, and can be used in gene therapy and vaccine production. The
CC ovarian tumour proteins and polynucleotides can be used to inhibit
CC the development of cancer, particularly ovarian cancer. They can also
CC be used to diagnose the onset and progression of cancer.
XX
SQ Sequence 580 BP; 155 A; 136 C; 172 G; 108 T; 9 other;

Query Match 22.4%; Score 412.2; DB 22; Length 580;
Best Local Similarity 90.0%; Pred. No. 7.6e-97;
Matches 441; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 234 gtcggcttttgaggaaacgggaaagaccacaaaggaagtgtgattgacacggctatggcatcct 293
Db 1 gtcagcctttgaggaaacggcgaagaccacaaaggaggtgtgattgacacggcgtatggcatcct 60
Qy 294 ggaagggaaggctctggagtcaggtacacacaaagtcggacttacggttaattgaagtcaac 353
Db 61 ggaccagaaggcctctggagtcaggtacacacaaagtcggacttgcggttaattgaagtcaac 120
Qy 354 tgagaccatttgcaagagctcttgactacagcctgcacaaaggagagactggcgagcaac 413
Db 121 tgagaccatttgcaagagctcttgattatagcctgcacaaaggagagaccggcgagcaac 180
Qy 414 ccggtttgcaagggtatgtcggagacotttgagacgtgtgcacaaacctagtcacaaagg 473
```

```
Db 181 tgcatttccaaggcgcgtgtcagagacctttgagacattacacaacctggtacacaaagg 240
Qy 474 ggtcaagtggtgatggtatccccctatgacctgtggaacgagacctcagcagaggtggc 533
Db 241 ggtcaagtggtgatggtacccccctatgacctgtggaacgagaccttcgagaggtggc 300
Qy 534 tgacctcaagaagcagtgctgacctgtggtggaagagtttgaagaggtgattgagagactg 593
Db 301 tgacctcaagaagcagtggtgacctgtggtggaagagtttgaagaggtgattgagagactg 360
Qy 594 gtacaggaaccaccaccaggaggagacctgactgaattcctctgtgcacaccacgtgtgaa 653
Db 361 gtacaggaaccaccaccaggaggagacctgactgaattcctctgtgcacaccacgtgtgaa 420
Qy 654 ggaagaagacacagattgctctacagacggtggtctgcaagaaggaggacacatagcctc 713
Db 421 gggaaaadacacacagttgctgcgcagagcagtggctccggcaagaaggaggagacacagacct 480
Qy 714 cctgggagg 723
Db 481 gcccgggcgg 490

RESULT 8
AAH42779
ID AAH42779 standard; cDNA; 274 BP.
XX
AC AAH42779;
XX
DT 21-AUG-2000 (first entry)
DE Human secreted expressed sequence tag SEQ ID NO:1519.
XX
KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
KW antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021990-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24205.
XX
PR 15-OCT-1998; 98US-0104435.
XX
PA (GEMV ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
DR WPI; 2000-317937/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 479; 618pp; English.
XX
CC AAH4261 to AAH43419 represent specifically claimed secreted expressed
```

CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnary; antiulcer; osteopathic; neuroprotective;
CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 274 BP; 76 A; 60 C; 81 G; 57 T; 0 other;

Query Match 12.2%; Score 224.4; DB 21; Length 274;
Best Local Similarity 91.9%; Pred. No. 2.6e-48; Mismatches 21; Indels 0; Gaps 0;
Matches 237; Conservative 0;
QY 307 tctggagtcgaagtcacacaaagtcggacttcggttaattgaagtcactgagaccatttcg 366
Db 5 tctggagtcgaagtcacacaaagtcggacttcggttaattgaagtcactgagaccatttcg 64
QY 367 aagagccttcgactacacgctgcacaaagagagactggcgacaccggtttgccaag 426
Db 65 aagagccttcgactacacgctgcacaaagagagactggcgacaccggtttgccaag 124
QY 427 ggtatgctcgagacctttgagcgtgcacaaactgctcacaaggggtcaaggtggtg 486
Db 125 ggtatgctcgagacctttgagcgtgcacaaactgctcacaaggggtcaaggtggtg 184
QY 487 atggatccctctatgagctgtggaacgagacacctcagagaggtggtgacctcaagaag 546
Db 185 atggatccctctatgagctgtggaacgagacacctcagagaggtggtgacctcaagaag 244
QY 547 cagtgagacgtgctggtg 564
Db 245 cagtgagacgtgctggtg 262

RESULT 9
AA249678
ID AA249678 standard; cDNA; 744 BP.
XX
AC AA249678;
XX
DT 07-APR-2000 (first entry)
XX Human myocardium secreted protein-5 coding region.
DE
XX Human; myocardium secreted protein-5; MSP-5; hypotensive; cardiac;
KW cardiac cellular process; cardiovascular disorder; cardiomyopathy;
KW congestive heart failure; hypertension; atherosclerosis;
KW coronary artery disease; valvular disease; arrhythmia; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..744
FT /*tag= a
FT /product= "MSP-5"
FT /note= "Myocardium secreted protein-5"

FT sig_peptide 1..66
FT /*tag= b
FT mat_peptide 67..744
FT /*tag= c
FT /label= Mature_MSP-5
XX WO967385-A1.
PN 29-DEC-1999.
PD 22-JUN-1999; 99WO-US13937.
XX 23-JUN-1998; 98US-0090398.
PR 29-SEP-1998; 98US-0163285.
XX (WILL-) MILLENNIUM PHARM INC.
PA Khodadoust MM;
PI WPI; 2000-136983/12.
XX P-PSDB; AAY44629.
DR Novel myocardium secreted protein-5 polynucleotides, used to modulate a
PT variety of cellular processes -
PT
XX
PS Claim 1; Pages 92-93; 99pp; English.
XX
CC The present sequence is the coding region of myocardium secreted protein
CC -5 (MSP-5) cDNA which was isolated from a cDNA library prepared from a
CC cardiac tissue sample obtained from a biopsy of a 42 year old woman
CC suffering from congestive heart failure. The MSP-5 protein has
CC hypotensive and cardiac activities. It is highly expressed in heart,
CC brain, placenta, foetal lung, liver, kidney, testis, small intestine and
CC pituitary gland. The present sequence is used to modulate a variety of
CC cellular processes, especially cardiac cellular processes. MSP-5 is used
CC to modulate the activity of one or more proteins involved in a
CC cardiovascular disorder, e.g. congestive heart failure or cardiomyopathy.
CC Conditions and diseases which can be treated include hypertension,
CC atherosclerosis, coronary artery spasm, coronary artery disease, valvular
CC disease, arrhythmias, and cardiopathies (e.g. hypertrophic, dilative, or
CC restrictive cardiomyopathies), and disorders related to under or over
CC expression of MSP-5.
XX
SQ Sequence 744 BP; 208 A; 143 C; 254 G; 139 T; 0 other;

Query Match 8.7%; Score 160.2; DB 21; Length 744;
Best Local Similarity 57.3%; Pred. No. 1.8e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
QY 162 ggctgagagagacgagctgggtgcgattgccagcaaatgcgaagtgtgcaagtattgttc 221
Db 75 ggaggagagcagatgacacagaacgcttgcgccagcaaatgcgaagtgtgtaagctgcgag 134
QY 222 tgtgagctgaagtcggcttttggagaaacgggaaagcaacgaagattgattgacacgg 281
Db 135 cacagactacagcggaactgagtcgcacggctcgatctcgagaggtgctcgagctggg 194
QY 282 ctatgcatctcgacgggaagggctctggagtcgaag---tacaccaagtcggactacg 338
Db 195 gcaggtgctgtacaggaaggaaggaagacacgcttcctacagcgttcacagagaag 254
QY 339 gtaattgagagtcactgagaccatttgcagagagctctgactacagcctgcacaagga 398
Db 255 gctggaagagcccttagaatttatgtgagcggtatctggtgactatagtgttcacgtga 314
QY 399 gaggactggcagcaacccggtttgccaagggtatgtcgagagacgttttgagcgtgcaca 458
Db 315 gcgaaggctcactgagatatgccaagggtcagatcagaccatggcaacactgaaagg 374
QY 459 cctagtccacaaggggtcgaaggtggtgtagtatccctatagctgtggaacagagac 518
Db 375 cctagtgcagaaggggtggaaggtggtatcctggtggatccctctgagctgtggatgagcc 434


```

FH Key Location/Qualifiers
FT CDS 108..854
FT /*tag= a
FT /transl_except= (pos: 423..425, aa: Lys)
FT sig_peptide 108..170
FT /*tag= b
PN WO200056889-A2.
XX 28-SEP-2000.
XX
XX 01-MAR-2000; 200WO-US05601.
XX
XX 23-MAR-1999; 99US-0125774.
XX 23-MAR-1999; 99US-0125778.
XX 24-MAR-1999; 99US-0125826.
XX 31-MAR-1999; 99US-0127035.
XX 05-APR-1999; 99US-0127706.
XX 21-APR-1999; 99US-0130359.
XX 27-APR-1999; 99US-0131270.
XX 27-APR-1999; 99US-0131291.
XX 04-MAY-1999; 99US-0132371.
XX 04-MAY-1999; 99US-0132379.
XX 04-MAY-1999; 99US-0132383.
XX 25-MAY-1999; 99US-0135750.
XX 08-JUN-1999; 99US-0138166.
XX 20-JUL-1999; 99US-0144791.
XX 03-AUG-1999; 99US-0146790.
XX 09-DEC-1999; 99US-0170262.
XX
XX (GETH ) GENENTECH INC.
XX
XX Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
XX Stewart TA, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI: 2000-628263/60.
XX P-PSDB: AAB18922.
XX
XX Novel secreted and transmembrane polypeptides useful for diagnosing
XX tumour in a mammal, for identifying agonists and antagonists of the
XX polypeptide and for therapeutic use
XX
XX Claim 2; Fig 27; 222pp; English.
XX
XX The present sequence encodes a secreted or transmembrane polypeptide.
XX The specification describes polypeptides designated PRO1484, PRO4334,
XX PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
XX PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
XX PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
XX useful for diagnosing tumour in a mammal. The polypeptides, their
XX agonists and antagonists are useful treating a condition associated with
XX expression or activity of the polypeptide. Conditions treated include
XX obesity, diabetes or hyper- or hypo-insulinemia. The polypeptides are
XX capable of inducing proliferation of mammalian kidney mesangial cells
XX and are therefore useful for treating kidney disorders associated with
XX decreased mesangial cell function such as Berger's disease or other
XX nephropathies associated with Schönlein-Henoch purpura, celiac disease,
XX dermatitis herpetiformis or Crohn's disease. The nucleic acids may be used
XX to generate transgenic animals for use in development and screening of
XX therapeutically useful reagents and also for chromosome identification
XX and tissue typing.
XX
XX Sequence 1470 BP; 373 A; 322 C; 438 G; 337 T; 0 other;

```

Search completed: May 24, 2002, 00:23:41
Job time: 9038 sec

```

Query Match 8.7%; Score 160.2; DB 21; Length 1470;
Best Local Similarity 57.3%; Pred. No. 2.4e-31;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

QY 162 ggctgagagaccgctgggtgctgattgccagcaaatgcgaagtgtgcaagtattgtgc 221
DB 182 ggagagagacgatgacacagacgcttcccccagcaaatgcgaagtgtgtaagctgtag 241

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 24, 2002, 00:12:35 ; Search time 3953.4 seconds
(without alignments)
9734.381 Million cell updates/sec

Title: US-09-696-686-47
Perfect score: 1839
Sequence: 1 ggcacgaggaggagcgcc.....taaaaaaaaaaaaaaaaaa 1839

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

RESULT 1
LOCUS BC013549
DEFINITION Mus musculus, RIKEN cDNA 1600025D17 gene, clone MGC:19340 IMAGE:4222133, mRNA, complete cds.
ACCESSION BC013549
VERSION BC013549.1 GI:15488835
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1882)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.

ALIGNMENTS

1	1829.4	99.5	1882	10	BC013549	BC013549 Mus muscu
2	1784.2	97.0	1909	10	AF361644	AF361644 Mus muscu
3	691.6	37.6	1422	9	BC008898	BC008898 Homo sapi
4	691.6	37.6	1422	9	BC008961	BC008961 Homo sapi
5	691.6	37.6	1720	9	BC004423	BC004423 Homo sapi
6	491.2	26.7	962	9	HSU80744	U80744 Homo sapien
7	412.2	22.4	580	6	AX198865	AX198865 Sequence
8	412.2	22.4	580	6	AX209392	AX209392 Sequence
9	331.2	18.0	913	9	BC022093	BC022093 Homo sapi
10	284	15.4	612	9	AF161347	AF161347 Homo sapi
11	186	10.1	89855	2	AC074220	AC074220 Mus muscu
12	160.2	8.7	745	6	AR142811	AR142811 Sequence
13	160.2	8.7	746	6	AX136616	AX136616 Sequence
14	160.2	8.7	747	6	BD006701	BD006701 Novel pol
15	160.2	8.7	1413	9	BC019903	BC019903 Homo sapi
16	160.2	8.7	1419	6	AX136423	AX136423 Sequence
17	160.2	8.7	1512	6	AR142810	AR142810 Sequence
18	156.6	8.5	113109	9	HS475N16	AL035587 Human DNA
19	153	8.3	1286	3	AY060987	AY060987 Drosophill
20	151.4	8.2	164999	3	AC009279	AC009279 Drosophill
21	151.4	8.2	279530	3	AE003518	AE003518 Drosophill
22	144.4	7.9	76187	2	AC015219	AC015219 Drosophill
23	84.8	4.6	129	11	G64272	G64272 17 Human Ho
24	75.2	4.1	7218	6	166494	166494 Sequence 14
25	67.8	3.7	100345	2	AC095760	AC095760 Rattus no
26	67.6	3.7	501	6	AX094199	AX094199 Sequence
27	67.6	3.7	510	6	AX079702	AX079702 Sequence
28	66	3.6	125020	9	AF429315	AF429315 Homo sapi
29	60.6	3.3	30925	2	AC107160	AC107160 Rattus no
30	60.2	3.3	125020	9	AF429315	AF429315 Homo sapi
31	60.2	3.3	198280	10	AL593843	AL593843 Mouse DNA
32	60.2	3.3	214801	2	AL596383	AL596383 Mus muscu
33	59.8	3.3	83175	2	AC098621	AC098621 Rattus no
34	59.8	3.3	140018	2	AC099090	AC099090 Rattus no
35	59.4	3.2	132764	9	AC073842	AC073842 Homo sapi
36	59.4	3.2	167854	2	AC092483	AC092483 Homo sapi
37	58.6	3.2	161334	10	AC091629	AC091629 Mus muscu
38	58.2	3.2	86430	2	AC100626	AC100626 Mus muscu
39	58	3.2	139608	2	AC103455	AC103455 Rattus no
40	58	3.2	186238	2	AC084828	AC084828 Mus muscu
41	58	3.2	198994	10	AL583886	AL583886 Mouse DNA
42	58	3.2	215914	2	AL603802	AL603802 Mus muscu
43	57.4	3.1	10732	6	E32986	E32986 Gene encodi
44	56.8	3.1	92904	2	AC107493	AC107493 Rattus no
45	56.6	3.1	97348	10	AF091216	AF091216 Mus muscu

cdDNA Library Preparation: Life Technologies, Inc.
cdDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 25 Row: 4 Column: 9.

FEATURES	SOURCE
----------	--------

CDS

	BASE COUNT	COUNT	ORIGIN
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	0	0	0
31	0	0	0
32	0	0	0
33	0	0	0
34	0	0	0
35	0	0	0
36	0	0	0
37	0	0	0
38	0	0	0
39	0	0	0
40	0	0	0
41	0	0	0
42	0	0	0
43	0	0	0
44	0	0	0
45	0	0	0
46	0	0	0
47	0	0	0
48	0	0	0
49	0	0	0
50	0	0	0
51	0	0	0
52	0	0	0
53	0	0	0
54	0	0	0
55	0	0	0
56	0	0	0
57	0	0	0
58	0	0	0
59	0	0	0
60	0	0	0
61	0	0	0
62	0	0	0
63	0	0	0
64	0	0	0
65	0	0	0
66	0	0	0
67	0	0	0
68	0	0	0
69	0	0	0
70	0	0	0
71	0	0	0
72	0	0	0
73	0	0	0
74	0	0	0
75	0	0	0
76	0	0	0
77	0	0	0
78	0	0	0
79	0	0	0
80	0	0	0
81	0	0	0
82	0	0	0
83	0	0	0
84	0	0	0
85	0	0	0
86	0	0	0
87	0	0	0
88	0	0	0
89	0	0	0
90	0	0	0
91	0	0	0
92	0	0	0
93	0	0	0
94	0	0	0
95	0	0	0
96	0	0	0
97	0	0	0
98	0	0	0
99	0	0	0
100	0	0	0

	Query Match	99.5%	Score 1829.4;	DB 10;	Length 1882;
	Best Local Similarity	99.9%;	Pred. No. 0;		
	Matches 1830; Conservative	0;	Mismatches	1;	Indels 0; Gaps
Qy	9	ggaggaacgcgcccgggtccgcgtctgtccttgggtcccgctggggccatggagtcacatgc	68		
Db	45	GGAGGAACGCCCGCGGCTCCGCTGTCTCTGGTTCGGCTGGGCATGGAGTCCATGTC	104		
Qy	69	tgaagtcgccccgcgtgcctctatttcttttgctgtgtgtcttccgctgctgtcctcct	128		
Db	105	TGAGTCTCGGCCCGCTGCCTTTATTTCCTTGCTGTGTCTGTCTTCGCTGCTGCTCCT	164		
Qy	129	tectgcccccgaagttagccgcgagtcgccgcgcccgggtgaggagaccgactgggtgcgatt	188		
Db	165	TCTTGCCCCCGAAGTAGGCCCGCGAGTCCC CGCGGGCTGAGGAGACCGACTGGTGCGGATT	224		
Qy	189	gcccagcaaatcgcgaagtygtcaaagta tttgtctgtgagctgaagtcggtctttgaaga	248		
Db	225	GCCCAGCAAAATCGAAGTGTGCAAGTATTTGCTGTGAGCTGAAGTCGGCTTTTCAGGA	284		
Qy	249	aacggaaagaccgaaggaagtgatgcacacggcttatggcatccctggacgggaagggctc	308		
Db	285	AACGGGAAGAACCAGGAAGTGATTTCACACCGGCTATGGCATCTTGGACGGGAAGGGCTC	344		
Qy	309	tggagtcaagtataccaagtcggaccttacggtttaattgaagtcactgagaccatttgaaa	368		
Db	345	TGGAGTCAAGTACACCAAGTCGGACTTACGGTTAATTGAAGTCAC TGAGACCATTTGGCAA	404		
Qy	369	sagccttctgactacagccttgcaagagagaggaactggcagcacccggtttgccaaggg	428		
Db	405	GAGGCTTCTGNACTACAGCTTGCAACAAGAGAGGACTGGCAGCACCCGGTTTGCCCAAGG	464		
Qy	429	tatgtcgagacctttgagcgtgcacaacctagtcaccaaggggttcaaggttgtat	488		
Db	465	TATGTCGGAGACCTTTGAGACGCTGCACAACCTAGTGTCCACAAGGGGTCAAGGTGGTAT	524		

|||||
Db 1605 AGTCTCCCTCATGATCTTTTGGCTTTGTCCCCCAATCCAGGGCTGGAAAGGCAT 1664
Qy 1629 caccatattgagcgttaacctctacttagaggtgctgaggcccgaggttg 1698
Db 1665 CACCATATTGGAGGCTTAACCTGTCTAGTTACTAGGAGGTGCTGGGAGCCCGGGGTG 1724
Qy 1689 gttggggtaatacactcactggctctcagccttctaactgacgccccttaatacagtt 1748
Db 1725 GTTTGGGGTAATCACTCACTGGCTCTCAGCCTTCTAACAACGTGACGCCCTTAATACAGTT 1784
Qy 1749 cttctctgtgtgtagctccacgccccccacacacacacacataaaattattcgatgctg 1808
Db 1785 CTTTCTGTGTGTGTGACTCCACGCCCCACACACACACACATAAAATATTTCGATGCTG 1844
Qy 1809 ttctaacactgtaaaaaaAAAAAAAAAAAA 1839
Db 1845 TTTCTAACTGTAATAAAAAAAAAAAAAA 1875
RESULT 2
AF361644 1909 bp mRNA linear ROD 12-MAY-2001
LOCUS Mus musculus putative retinoic acid-regulated protein mRNA,
DEFINITION complete cds.
ACCESSION AF361644
VERSION AF361644.1 GI:14028766
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1909)
AUTHORS Glozak,M.A., Li,Y., Reuille,R. and Rogers,M.B.
TITLE Trapping and characterization of novel retinoic acid response
elements
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1909)
AUTHORS Glozak,M.A., Li,Y., Reuille,R. and Rogers,M.B.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2001) Biology, University of South Florida, 4202
E. Fowler Avenue, Tampa, FL 33620, USA
FEATURES
source Location/Qualifiers
1..1909
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="P19; derived from matings between C3H/He
females and male animals carrying an X chromosome derived
from a feral mouse (McBurney & Rogers, Dev. Biol., 89,
503)"
/cell_type="embryonal carcinoma"
/clone_lib="24 hr retinoic acid-treated P19 cDNA (Bouillet
et al., Dev. Biol., 170, 420)."
89..919
/codon_start=1
/product="putative retinoic acid-regulated protein"
/protein_id="AAK52494.1"
/db_xref="GI:14028767"
/translation="MESMSELAPRCLLFPILLLLPILLPLLPAPKLGPSPAGAEEDWVR
LPKVECKTVAVELKSAFEETKTEVIDTGYILDGKSGVYKTKSDLRLIEVET
IKRLLDYSLHKRTSGNSRFKGMSETFETLHNLVHKGVKVMVDIPYELWNETSAEVA
DLKKQCDVLEEFEEVIDWRNHQBEDLTFELCANHLVKDKTSCIAERWSGKGD
ASLUGKKKKRSGVKGSSSSSKQKELGGLGEDANAEIEEGVQKASPLPHSPDPDEL
"
BASE COUNT 424 a 536 c 558 g 386 t 5 others
ORIGIN
Query Match 97.0%; Score 1784.2; DB 10; Length 1909;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
Qy 9 ggaggagcgcgcgggtcgtctgctggtccggtggccatggagtcacatgctc 68
|||||

|||||
Db 43 GGAGGAAGCCCGCGGTCCGCTCTGCTCTGGTCCGGCTGGGCCATGGATCCATGTC 102
Qy 69 tgagctcgcccccgcgtgctcttattcttctgctgctgctcttccgctgctgctcct 128
Db 103 TGAGCTCGCCCGCGGTCTTATTTCCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 162
Qy 129 tcttgcctccgaagctagggcccgagtcctcccggggctgagagagccgagctgggtgagatt 188
Db 163 TCTTGCCTCCGAGCTAGGCCCGAGTCCCGCGGGGCTGAGGAGACCCACTGGGTGCCATT 222
Qy 189 gccagcaaatggaagcgtgcaagtatgttctgtgtgagctgaagtcggtcttttgaggga 248
Db 223 GCCCAGCAATGCCAAGTGTCAAGTATGTGTGTGTGAGCTGAAGTCGGCTTTTGAGGA 282
Qy 249 aacgggaagacacaaaggaagtattgacacggctatgagcatcctgagcggggaagggctc 308
Db 283 AACGGAAAGACCAAGGAAGTGTATGACACCGGCTATGGCATCTCGACGGGAAGGGCTC 342
Qy 309 tggagtcgaagtacacaaagtcggacttcaggttaattgaagtcactgagaccattgcaa 368
Db 343 TGGAGTCAAGTACACCAAGTCGGACTTACCGTTTAATTTGAAGTCACTGAGACCATTTGCAA 402
Qy 369 gaggtcttctgactacagcctgcacaagagagagactggcagcaaccggtttgccaaggg 428
Db 403 GAGGCTTCTGACTACAGCCTGCACAAGGAGAGACTGGCAGCAACCGGTTTGGCAAGGG 462
Qy 429 tatgtggagaccttgagacgctgcacaacctagtcacaaaggggtcacaaggtcgaagtgat 488
Db 463 TATGTGCGAGACCTTTGAGACGCTGCACAACCTAGTGTGAGGACTGGTACAGGAACCACTCA 522
Qy 489 ggatccccctatgagctgtggaacgagacctcagcagaggtggtgacctcagaagaca 548
Db 523 GGATATCCCTATGAGCTGTGGAAACGAGACCTCAGCAGAGGTGGCTGACCTCAAGAAGACA 582
Qy 549 gtgtgacgtgctggtggaagagtttgaagaggtgattgagactggtcacaggaaaccacca 608
Db 583 GTGTGACGTGCTGGTGGAGAGTTTGAAGAGGTGATTGAGGACTGGTACAGGAACCACTCA 642
Qy 609 ggaggagaacctgactgaatactctctgtgccaaacacgctgctgaaggaagagacagag 668
Db 643 GGAGGAAGACCTGACTCAATTCCTCTGTGCCAACACACGCTGTGAAGGGAAGGACACGAG 702
Qy 669 ttgcctagcagagcgtgtgcttggaagaaggggagacatagcctccctggagggaagaa 728
Db 703 TTGCCTAGCAGAGCGGTGGTCTGGCAAGAAGGGGAGACATAGCCTCCCTGGGAGGGAACAA 762
Qy 729 atccaagaagaagcgcagcgaggttcaagggtcctcctcagtgagcagcagcagcagagaa 788
Db 763 ATCCAAGAAGAAGCGCAGCGAGTCAAGGGCTCTCTCCAGTGGCAGCAGCAAGCAGAGGAA 822
Qy 789 ggaactggggggcctgggggagagatgccaaacgagagagagagaggggtgtgcagaagggc 848
Db 823 GGAACCTGGGGGGCTGGGGGAGGATGCCAACCGCGAGGAGGAGGCTGTGCAGAAGGC 882
Qy 849 atcgccctccacacagacccccctgagctgtagctgtagccagcttagtgtccttgaatc 908
Db 883 ATGCCCTCTCCACACAGCCCTCTGATGAGCTGTGAGCCACAGCTTAAGTCTCTTGAATC 942
Qy 909 aagaccctgacttcagagcttggaacacacacagcgagcagcgagcagctccagcaag 968
Db 943 AAGACCCTGACTTACAGAGTTGGGACACGCACA -CGCAGCGCAGCGCAGCTCCAGCAAG 1001
Qy 969 gacagctgctgccagcagcagctcctcctcctggtgtgccccttctcctcctgaa 1028
Db 1002 GACAGCTGCTGCCAGCATCAGGTCTCTCTCTCTGGCTGTGCCCTTTCTCTCTCTTGA 1061
Qy 1029 caacagcaagaggtggaagatctgggggtgctgggagacggcaccacaaagggaagagga 1088
Db 1062 CAACAGCAAGAGGTGGAAGGATCTGGGGTGTGGGAGACCGGCACCCCAAGGAGGAGAGA 1121
Qy 1089 ggaggagcagaagcagctctcttctacacagtcctccctcagcagctccgggtccacc 1148
|||||

D	b	1122	GGAGGACAGAGGAGCTCTCTTTCTACAGTCCCTCCCTCAGAGCTCCGGGGTCCACC	1181
Q	y	1149	cagcatccccaggctgagatccaggtcctgacatggaagctgaagcatgagcacat	1208
D	b	1182	CAGCATCCCAGGCTGAGATCCAGGCTCTGACATGGAAGCTGAAGACATGAGGCACAT	1241
Q	y	1209	aagtgtccacagcgcccccttaagccagaaaggactcgtgagctcagcagccaagg	1268
D	b	1242	AAGATGTCTACAGCGGCCCTTCAGCCAGAAAGGACTCCGTGACGCTCAGCAGCCAGG	1301
Q	y	1269	cctgcctctcttcaccaaagcattctctctgctggtcctgtcgagatggtaaattcg	1328
D	b	1302	CCTGCCTCTTCTCCACCAGCATTCCTCTCTGCTGCTCTGTCGGATGGTAAATTCG	1361
Q	y	1329	agaacttcaggacaaactcgggtgtgtgcacaaaggggctggagccagagccagagcca	1388
D	b	1362	AGAACTTCCAGGACAAACTCGGGGTGTGGCAAAAGGGGCTGGAGCCAGACCCAGAGCA	1421
Q	y	1389	gcagagactgcagagagggcacctgacctaaacccccctgaaagccaatctgcagttc	1448
D	b	1422	CGCCAGAGACTGCAGAGAGGGCACCCTGACCTAACCCCTCGAAAGCCAACTCGCAGTTC	1481
Q	y	1449	ccgtgtccacccactctcctgtgaggaagcctcattgtctgtccagagcccttcccagggc	1508
D	b	1482	CCGTGTCCACCACACTCCTCTCTGAGACGCCCTCATGTCTGCCTCCAGCCCTTCTCCAGGCG	1541
Q	y	1509	taccagagtaaacaccttttggcctttgcgttttggttctcgtggtcctcatcagcctccag	1568
D	b	1542	TACCAGAGTAACACACTTTTGCCCTTTCCGGTTTGGTTCTGGTCTCTCATCAGCCTCCAG	1601
Q	y	1569	agtgccctcatcgatcttttttgcctttgtcccccataccaggggctggaaggccat	1628
D	b	1602	AGTGTCCCTCATCGATCTTTTGTGCTTTGTCCCCATCCACGGGTGNTGGAAGGCCAT	1661
Q	y	1629	cacatcattggaggcttaacctgtcagttactaggagtgctgggagcgccgggggttg	1688
D	b	1662	CACCATCAATTCNAGGNTTAACTGTGTCAGTTACTAGAAGGTGCTGGGAGCGCCCGGGGTG	1721
Q	y	1689	gttggggtaactcactcactggctctcagcttctaaactgcagcccttaatacagtt	1748
D	b	1722	GTTTGGGGTAATCACTNACTGCCTCAGCCTTANAACACTGCAGCCCTTAATACAGTT	1781
Q	y	1749	cctctgtgtgtgtagctccacgccccacacacacacataaaattatttcgatgctg	1808
D	b	1782	CTTTCTGTGTGGTGACTCCACGCCCCACACACACACACATAAAATTTATTGGATGCTG	1841
Q	y	1809	tttcataactgtaa	1822
D	b	1842	TTTTCAATACTGTAA	1855
RESULT	3			
BC008898				
LOCUS			1422 bp	mRNA
DEFINITION				linear
ACCESSION				PRI 12-JUL-2001
VERSION				IMAGE:2959532, mRNA, complete cds.
KEYWORDS				
SOURCE				
ORGANISM				
				Homo sapiens
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
				Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
				1 (bases 1 to 1422)
				Strausberg, R.
				Direct Submission
				Submitted (29-MAY-2001) National Institutes of Health, Mammalian
				Gene Collection (MGC), Cancer Genomics Office, National Cancer
				Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
				USA
				NH-MGC Project URL: http://mgc.nci.nih.gov
				Contact: MGC help desk
				Email: cgabs-r@mail.nih.gov
				Tissue Procurement: ATCC
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REMARK				
COMMENT				

CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate; 2 Row; m Column: 12
 Best Local Similarity 84.5%; Pred. No. 4.8e-164;
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6002306.

```

FEATURES             Location/Qualifiers
     source            1..1422
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="MGC:1220 IMAGE:2959532"
                        /tissue_type="Colon, adenocarcinoma"
                        /clone_lib="NIH_MGC_15"
                        /lab_host="DH10B-R"
                        /notes="Vector: pOTB7"
                        78..914
     CDS               /codon_start=1
                        /product="Unknown (protein for MGC:1220)"
                        /protein_id="AAH08898.1"
                        /db_xref="GI:114286204"
                        /translation="MDSMPEPASRCKLLPLLLLLLLLPAPELGPSSQAGAEENDWVR
                        LPKCBVCYKVAVELKSAFETGKTKEVIGTGILDQKASGVKYTKSDLRLLIEVTET
                        IKRLLDYSLHKRETCNSRFKMGSTFTFLHNLVHKGVKVMIDIPVELWNETSAEVA
                        DLKKDDLVVEEFEEIWDYRNHOEDLTFFLCANHLVKGDTSCLAEOWSKKGDT
                        AALGGCKSLKKSRAAKGRSSSRQRLEGLGDPSPDEEGTQKASPLTHSPHPD
                        EL"
BASE COUNT           351 a   400 c   422 g   249 t
ORIGIN
Query Match          37.6%; Score 691.6; DB 9; Length 1422;
Best Local Similarity 84.5%; Pred. No. 4.8e-164;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

QY    5 cgaggaggaagccgcgggctcgtcgtcgtcggttcggtccgctgggccatggagtcca 64
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    28 CGCGGAGGAGGAACGCCGGCTCCTTAGGTCCGGGCCGGCGGCATGATTCAA 87

QY    65 tgtctgagctgcgcgcgcgcgtcgtctatttcctttgctgctgctctccgctgtgc 124
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    88 TGCTGAGCCGGCTCCGGCTGTTCTTGCTTCTTCCCTTGTCTGCTGCTGCTGCTG 147

QY    125 tccttcctgcgcgaactaagcccagtcacgcgcgggctgaagagaccagactgggtgc 184
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    148 TGTCTCCGGCCCCGGAGCTGGCGCCGAGCACAGGCCGAGTGTGAGGAGAACCAC 207

QY    185 gattgccagcaaatgcgaagtgtgcaagtatgttgctggagctgaagtgcgctttg 244
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    208 GCCTGCCCAAGAAATGGAAGTGTGTAATATGTTGCTGGAGCTGAAGTCACCC 267

QY    245 aggaacaggaaacacaaaggaagtgttacacccgctatggcatacctggacgggaag 304
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    268 AGGAACCGCAACACCAAGGAGTGATTGSCACGGGTATGGCATCCTGGACCA 327

QY    305 gctctgagtcgaatacaccaagtgcgaccttaattgaagtcaactgagaccattt 364
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    328 CCTCTGGAGTCAAAATACACCAAGTCGGACTTGGCGTTAATCGAAGTCACTG 387

QY    365 qcaagagctcttgacctacagctgcacagagagagactggcagaacccggttgtcca 424
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    388 GCAAGAGCTCCTGGATTATAGCTTGCAACAGAGAGAGCGCGCAGCAATCGATT 447

QY    425 aggtatgtcggagacctttgagacgctgcacaacaccttagtcacaaaaggggtc 484
      ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    448 AGGCATGTCAGACACTTTTCAGACATATACACCTGGTATACACAAAGGGGTCAAG 507

```


Db 748 AGAAGTCCAAAGAGAGAGACAGCAGCGCCAGCCAGCACCGCGGAGGAGTAGCAGCAGCA 807

QY 779 agcagaggaagaactgggggctgggggagagatgcccaaccgagggagggagggg 838

Db 808 AACAAAGGAAGGAGCTGGGTGGCTTGTAGGAGACCCAGCCCGGAGGAGTAGGGCA 867

QY 839 tcagagaagcatcgccctccacacagcccccctgatgagctgtgagccagcttagtg 898

Db 868 TCAGAAAGCATCCCTCTCACACAGACGCCCTGATGAGCTCTGAGCCACCCAGCAT 927

QY 899 tcttgaatcaagacccctgacttcagagcttgga 934

Db 928 CTCCTGTCTGAGACCCCTGATTTGAAGCTGAGGA 963

RESULT 5

BC004423

LOCUS BC004423 1720 bp mRNA linear PRI 12-JUL-2001

DEFINITION Homo sapiens, clone MGC:3530 IMAGE:2819660, mRNA, complete cds.

ACCESSION BC004423

VERSION BC004423.1 GI:13325207

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1720)

Direct Submission

Strausberg, R.

Submitted (12-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-re@mail.nih.gov

Tissue Procurement: DCTD/Dnp

CDNA Library Preparation: Rubin Laboratory

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

BC Cancer Agency; Genome Sequence Centre, info@bcqsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 9 Row: 0 Column: 7

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5729791.

Location/Qualifiers

1. .1720

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:3530 IMAGE:2819660"

/tissue_type="lung, small cell carcinoma"

/clone_lib="NIH_MGC_7"

/lab_host="DH10B-R"

/note="vector: pOTB7"

388. .1224

/codon_start=1

/product="Unknown (protein for MGC:3530)"

/protein_id="AAH04423.1"

/db_xref="gi:13325208"

translation="MDSMPAPSRCLLLPLLLLLLLPAPELGPSQAGAEWDWR LPSRCEVKYVAVELKAFETGKTKEVIGTGYILDQKASGVKTKSLRLIEVTIC KRLDYSLHKERTGSRNRFKGMSETFETLHNLVHKGVKVMVDIPVELMNETSAEVA

FEATURES

source

CDS

DLKKOCDVLVEEFVIEDWYRNHQBEDLTPELCANHLVKGKDTSCSLABQWSKKGDT
AALGGKSKKKSSRAKAAGRSSSKQRKELGLEDGSPDEEDGIQKASPLTHSPPD
EL"

BASE COUNT 389 a 491 c 526 g 314 t

ORIGIN

Query Match 37.6%; Score 691.6; DB 9; Length 1720;
Best Local Similarity 84.5%; Pred. No. 4.8e-164;
Matches 791; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

QY 5 cgagaggaggaagcgccggtccgctgtctgtggtccgggtcggtggcgaatgaatcca 64

Db 338 CGCGGGAGGAGGAACCGCCCGGTCTTTAGGGTCCGGCCCGCGCCGCGCATGGATTCAA 397

QY 65 tgtctgagctcgcccccgctcctcttatttcttctgtgtgtgtgtgtgtgtgtgtgtgt 124

Db 398 TGCTGTGAGCCCGGTCCCGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 457

QY 125 tcttctgtccccgaagctagtagcccgagtcgcccgcggtgtgtgtgtgtgtgtgtgtgt 184

Db 458 TGCTGCGCGCCCGGAGCTGGGCGCGGAGCCAGCGCGGAGCTGAGGAGAAACGACTGGGTTTC 517

QY 185 gattgcccaagaatcggaagtgtgcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 244

Db 518 GCTGTGCGCCAGCAATGCGAAGTGTGTAAATATGTCTGTGTGGAGCTGAAGTCAAGCTTTG 577

QY 245 aggaacggggaagacccaaggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 304

Db 578 AGGAACCGGCAAGACCAAGAGGTGATTGGCAGCGGCTATGGCATCTTGACCAAGG 637

QY 305 gctctggagtcgaagtacacaaagtcggacttacctgtgtgtgtgtgtgtgtgtgtgtgt 364

Db 638 CCTCTGGAGTCAATACACCAAGTCGGACTTTCGGTCTTAATCGAAGTCACTGAGACCATTT 697

QY 365 caaagaggtcttgactacacccctgcacaaagagagagagagagagagagagagagagag 424

Db 698 GCAAGAGGTCTCTGGATTATGCTTCACAGAGAGAGACCGCGCAGCAATTCGATTTCGCA 757

QY 425 aggggtatgtcggagacaccttggagcgtgcacaaacctagtcacacaaaggggtcaaggttg 484

Db 758 AGGGCATGTTCAGAGACCTTTTGAGACATTACAAACCTGTACACAAAGGGGTCAAGGTGG 817

QY 485 tgatgatatacccttatgactgt 544

Db 818 TGATGGACATCCCTATGAGCTGTGGAAGCAGACTTCTGCAGAGGTGGCTGACCTCAAGA 877

QY 545 agcagtgtgactgtctgt 604

Db 878 ACCAGTGTGATGTCTGT 937

QY 605 accagaggagaacctgactgaattctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 664

Db 938 ACCAGGAGGAAGACCTGTACTGAATTCCTCTGCGCCAAACCACTGTGTGAAGGGAAGACA 997

QY 665 cgagtgtctagcagagcgt 724

Db 998 CCAGTTGCTGTGCAGAGCAGT 1057

QY 725 aqaatccaagaagaagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 778

Db 1058 AGAAGTCCAAAGAAGAGACAGCAGCGGCCAAGGCAGCAGCGCGGAGGAGTAGCAGCAGCA 1117

QY 779 agcagagaagaagaactggggggt 838

Db 1118 AACAAAGGAAGGAGCTGGGTGGCTTGTAGGAGAGACCCAGCCCGGAGGAGTAGAGGCA 1177

QY 839 tcagagaagcatcgccctccacacagcccccctgatgagctgtgtgtgtgtgtgtgtgtgtgt 898

Db 1178 TCAGAAAGCATCCCTCTCACACAGACGCCCTTGTGATGAGCTCTGAGCTCAGCCCAAGCAT 1237

QY 899 tcttgaatcaagacccctgacttcagagcttgga 934


```
Db 241 GGTCAAGTGGTGATGGACATCCCTATGAGCTGTGGAACGAGACTTTCGACAGGTGC 300
QY 534 tgacctcaagaagcagtgtagctgctgtggaagagtttggaagtgattgagactg 593
Db 301 TGACCTCAAGAAGACAGTGATGCTGTGTGTGGAAGAGTTTGAGGAGTGATCGAGGACTG 360
QY 594 gtacaggaaccaccaggaggaagacctgactgaatttcctctgtgccaaaccacgtgctgaa 653
Db 361 GTACAGGAACACCAGGAGGAAGACCTGACTGAATTCCTCTGCGCCAACACCTGCTGAA 420
QY 654 gggaaaggacacagtgctgctgacagagcggtggtgtggaagaaggaggacatagcctc 713
Db 421 GGGAAAGACACAGTGTGCTGTCGACAGCANTGGTCCGCAAGAAGGGAGACACAGCCT 480
QY 714 cctgggaggg 723
Db 481 GCCCGGGCGG 490

RESULT 8
LOCUS AX209392 580 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 1232 from Patent WO0157207.
ACCESSION AX209392
VERSION AX209392.1 GI:15423815
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Algate,P.A. and Mannion,J.
TITLE Compositions and methods for the therapy and diagnosis of ovarian
cancer.
JOURNAL Patent: WO 0157207-A 1232 09-AUG-2001;
CORIXA CORPORATION (US)
FEATURES
Source i. .580
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 155 a 136 c 172 g 108 t 9 others
ORIGIN
Query Match 22.4%; Score 412.2; DB 6; Length 580;
Best Local Similarity 90.0%; Pred. No. 2.4e-93;
Matches 441; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 234 gtccgcttttgaggaacacgggaaagaccaggaagtgttgacacccgctatggctcct 293
Db 1 GTCAGCCTTTGAGGAACCGCAAGCAAGGAGGTGATTGACACGGGCTATGGCATCCT 60
QY 294 ggaagggaaggctctggagtcaggtacacccaagtcgacttcaggttaattgaagtac 353
Db 61 GGACCAGAAGGCTCTGGAGTCAATATACCAAGTCGGACTTTCGGGTTTATCGAAGTCAC 120
QY 354 tgaacacattgcaagaggctcttgactacacgcctgcacaaaggagagactggcagcaa 413
Db 121 TGAGACCAATTGTGACAGGCTCTCTGGATTATAGCTTGCAACAGGAGGACCGGACGAA 180
QY 414 ccggtttgcaagggttatgtcgagaccctttgagacgctgcacaaacctgctcacaaagg 473
Db 181 TCGATTGTGCAAGGGCATGTCAGAGACCTTTGAGACATTACAACTCGTACACAAAGG 240
QY 474 ggtcaagggtgatgatattccctatgactgtggaacgagacctcagcagagtgcc 533
Db 241 GGTCAAGGTGGTGATGGACATCCCTTATGAGCTGTGGAACGAGACTTTCGACAGGTGCC 300
QY 534 tgacctcaagaagcagtgtagctgctgtggaagagtttggaagtgattgagagactg 593
Db 301 TGACCTCAAGAAGACAGTGATGCTGTGTGGAAGAGTTTGAGGAGGTGATCGAGGACTG 360
QY 594 gtacaggaaccaccaggaggaagacctgactgaatttcctctgtgccaaaccacgtgctgaa 653
```

```
Db 361 GTACAGGACCAACGAGGAGAGACCTGACTGAATTCCTCTCGCCAACCACTGCTGAA 420
QY 654 ggaagaagacagagttgcctagcagcggtggtctgcaagaaggaggacatagcctc 713
Db 421 GGGAAAGACACAGTGTGCTGTCGACAGCANTGGTCCGCAAGAAGGGAGACACAGCCT 480
QY 714 cctgggaggg 723
Db 481 GCCCGGGCGG 490

RESULT 9
LOCUS BC022093 913 bp mRNA linear PRI 24-JAN-2002
DEFINITION Homo sapiens, clone IMAGE:3534003, mRNA, partial cds.
ACCESSION BC022093
VERSION BC022093.1 GI:18314384
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Strausberg,R.
TITLE Direct Submision
JOURNAL Submitted (22-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue procurement: DCTD/DTp
CDNA Library preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 11 Row: h Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
source Location/Qualifiers
i. .913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3534003"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
<1. .422
/codon_start=3
/product="Unknown (protein for IMAGE:3534003)"
/protein_id="AAH22093.1"
/db_xref="GI:18314385"
/translation="HEEYVMDIPYELWNETSAEVADLKQCDVLYVEEFEEVIEDWYRN
HQEDLTFLCANHVILKGDTSCLAEOWSGRKGDTAALGGKKKKKKRAAAGRSS
SSKORKELGGLEGDPSPDEDEGIQKASPLTHSPDEL"
BASE COUNT 228 a 263 c 268 g 154 t
ORIGIN
Query Match 18.0%; Score 331.2; DB 9; Length 913;
Best Local Similarity 84.0%; Pred. No. 7.6e-73;
Matches 388; Conservative 0; Mismatches 68; Indels 6; Gaps 1;
```

Qy	479	aggtggtgatggataatccccctatgagctgtggaacagagacctcagacagaagttggtgcacc	538
Db	10	AGTGTGTTATGGACATCCCCCTATGAGCTGTGGAAACGAGACTTCTGCAGAGGTGGCTGACC	69
Qy	539	tcaagaagcagtgtagcgctgctgagtgaagagtttgaagagggtgattgagagactggtaca	598
Db	70	TCAAGAAGCACTGTGATGTCTCGTGGAAAGATTGTGAGGAGGTGATCCAGGACTTGTTACA	129
Qy	599	ggaaccaccaggaggaagacctgactgaaattctcttgtccaaccacgctgctaagaaggaa	658
Db	130	CGAACCAACAGGAAGAGACCTGACTGAATTCCTCTCGGCCAACACGCTGCTGAAGGGAA	189
Qy	659	aggacacaggttccttagcagaagcgggtggtctcgtccagaagggggacatatagctccctgg	718
Db	190	AAGACACCACTTCCTCGCAGACAGCTGGTCCGCCAAGAAGGAGAGACACAGCTGCCCTGG	249
Qy	719	gagyggaagaaatccaagaagaagcgcagcgaggtcaaag-----ggctctctccagtgcca	772
Db	250	GAGGGAAGAAGTCCMAGAAAGAGCAGCAGGCGCAAGGCAGCAGGCGCAGGAGTAGCA	309
Qy	773	gcagcaagcagaagaagaaactgggggggctctgggggaggaatgccaacgcgcgaagagaggg	832
Db	310	GCAGCAAACAAGAGGAAGGAGCTGGTGTGGCTTTCAGGGAGAGCCCCAGCCCAGAGSGAGTG	369
Qy	833	agggtgtgcagaaggaatgcctcccacacacagccccctctgatgagctgtgagcccacgc	892
Db	370	AGGGCATCGAAGAGGCATCCCCCTTCACACACAGCCCCCCCTGATGAGCTGTGAGCCACC	429
Qy	893	ttagtgctcttgaatcaagaccctcgtactcaagagcttggga	934
Db	430	CAGCATCCTCTGTCTCGAGACCCCTGATTTTGAAGCTTGAGGA	471

RESULT 10

AF161347	AF161347	Homo sapiens HSPC084 mRNA, partial cds.	612 bp	mRNA	linear	PRI 01-FEB-2000
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						

FEATURES SOURCE

```

1: 1042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CBDABell"
/cell_type="cd34+ stem cell"
/tissue_type="cord blood"
<1..431
CDS

```

CDs

```

/codon_start=3
/product="HSPC084"
/protein_id="AAF28907.1"
/db_xref="GI:6841108"
/translaton="NETSRAVDL
HVLKGDTSCLAEQWSGKKGDTRAA
TPAPRRMRSRRPLSHPTAPLMSS
201 a 151 c 169 g
BASE COUNT

```

Query Match	15.4%	Score 284;	DB 9;	Length 612;
Best Local Similarity	86.7%;	Pred. No. 6.6e-61;		
Matches 338;	Conservative 0;	Mismatches 45;	Indels 7;	Gaps 2;
Qy 509	ggaacgagacctcagcagaggtggctgactcaagaagcagctgtgcgctgtgctgtggaag	568		
Db 1	GGACGAGACTTCTGCAGAGGTGGCTCAACCTCAGAGACGAGTGTGATGTGCTGGTGGGAAG	60		
Qy 569	agtttgaagaggtgatgtgaggaactgtgtacaggaaccaccaggaggaagaactgactgaat	628		
Db 61	AGTTTGAAGAGGTTATCCAGGACTTGGTACAGGAACACCACGAGGAGGAAGACTGACTGAAT	120		
Qy 629	tcctctctgtccaacccagctctctaaaggaagacacagagttgtctagcagagcgttggt	688		
Db 121	TCCTCTCTGGCCNACACAGTGTCTGAAGGGAAGACACAGTGTGCTGGCAGACAGTGGT	180		
Qy 689	ctggcaagaagggggacatagctccctgggaggggaagaaatcccaagaagaagcgacgcg	748		
Db 181	CCGGCAAGAAGGGAGACACAGCTGCCCTGGGAGGGGAAGAGTCCCAAGAAGAAGACGACGA	240		
Qy 749	gagtcaag-----ggctctctccagtgtggcagcagcaagcagaggaagaaactggggggcc	802		
Db 241	GGGCCAAGGCAAGCAGGCGGCAAGGAGTAGCAGCAGCAACAAGGAAGAGGAGCTGGGTGGCT	300		
Qy 803	tggggaggatgtccaacgcgcagagagagaggggtgtgcagaagggcatcgccccctcccc	862		
Db 301	TGAGGGA-GACCCCAAGCCCGAGGAGGATGAGGGCATCCAGAAAGCATCCCTCTCACAC	359		
Qy 863	acagcccccttgatgactgtgagccacgc	892		
Db 360	ACAGCCCCCTTGATGAGCTCTGAGCCCCACC	389		

RESULT 11

AC074220/C	AC074220	89855 bp	DNA	linear	HTG 25-JAN-2002
LOCUS	Mus musculus chromosome 17 clone RP23-76116 strain C57BL6/J,				
DEFINITION	WORKING DRAFT SEQUENCE, 38 unordered pieces.				
ACCESSION	AC074220				
VERSION	AC074220.7	GI:11876390			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	1 (bases 1 to 89855)				
	Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,				
	Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,				
	Gordon, M., Goltz, J.S. and Kucherlapati, R.				
TITLE	High Throughput Mouse Sequencing				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 89855)				
AUTHORS	Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,				
	Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,				
	Gordon, M., Goltz, J.S. and Kucherlapati, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2000) Department of Molecular Genetics, Albert				
	Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,				
	Bronx, NY 10461, USA				
	On Apr 28, 2001 this sequence version replaced gi:11276104.				
COMMENT	-----Genome Center				
	Center: Harvard Partners Genome Center				
	Center Code: HPGC				
	Web site:				
	http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/sequence/mouse.html				
	Contact: gnktm@capecod.bwh.harvard.edu				
	-----Summary Statistics				
	Center project name: ADY				
	Sequencing vector: pUC18; L08752				
	Chemistry: Dye-terminator Big Dye; 100%				

misc_feature /note="assembly_name:Contig122"
70116..71821
misc_feature /note="assembly_name:Contig121"
71842..73506
misc_feature /note="assembly_name:Contig120"
73527..74749
misc_feature /note="assembly_name:Contig119"
74770..75963
misc_feature /note="assembly_name:Contig118"
75984..77905
misc_feature /note="assembly_name:Contig117"
77926..79236
misc_feature /note="assembly_name:Contig116"
79257..81332
misc_feature /note="assembly_name:Contig115"
81353..83169
misc_feature /note="assembly_name:Contig114"
83190..84096

Query Match 10.1%; Score 186; DB 2; Length 89855;
Best Local Similarity 99.5%; Pred. No. 4.4e-36;
Matches 197; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 9 ggaggaagcgcgcgggtccgctctgtctgtgggtccggtgggccaatggagtcattgc 68
Dy 6070 GGAGGAAGCGCGCGGCTCGCTCTGCTCTGGGTCCGGCTGGGCAATGGAGTCCATGTC 6011
Qy 69 tgaactgcgcgcgcgcgcctctattctcttctgtgc-tgctgtctccgcgtgcctcc 127
Dy 6010 TGAGCTCGCGCGCGCTGCTCTATTTCCTTGTGCTGCTTCCGCTGCTGCTGCC 5951
Qy 128 ttctgtcccggaagctagggccgagtcgccgcggggtgagagaccgactgggtgcgat 187
Dy 5950 TTCCTGCCCGCAAGCTAGCGCGAGTCCCGCGGGGCTGAGGAGACCGACTGGGTGCGAT 5891
Qy 188 tggccagcaaatgcgaag 205
Dy 5890 TGCCGAGCAATGGCAAG 5873

RESULT 12

AR142811 AR142811 744 bp DNA linear PAT 08-AUG-2001
LOCUS Sequence 3 from patent US 6204013.
DEFINITION
ACCESSION AR142811
VERSION AR142811.1 GI:15104097

KEYWORDS
SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 744)

AUTHORS Khodadoust,M.M.

TITLE MSP-5 nucleic acid molecules and uses therefor

JOURNAL Patent: US 6204013-A 3 20-MAR-2001;

FEATURES Location/Qualifiers

1..744

source /organism="unknown"

BASE COUNT 208 a 143 c 254 g 139 t

ORIGIN

Query Match 8.7%; Score 160.2; DB 6; Length 744;
Best Local Similarity 57.3%; Pred. No. 1.4e-29;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

Qy 162 ggctgaggagaccgactgggtgcgattgccagcaaatgcgaagtgtgcaagtattgtgc 221
Dy 75 GGAGGAGCAGCATGACACAGAACGGCTTGCACGAGCAATGCCGAAGGTCTAAAGCTGCTGAG 134
Qy 222 tgtgagctgaagtgcgctttttgagaaacggaagaccgaaggaagtgtgacaccgg 281
Dy 135 CACAGAGTACAGGGCGGAAGTGTGAGTCCGACCGTGCATCTCGAGAGGTGCTGGAGCTGGG 194

Qy 282 ctatggcatcctgcacgggaagggtcctggagtcaag---tacaccaagtcggacttacg 338
Dy 195 GCAGGTGCTGGATACAGGCAAGAGGAGACACACGTGCTTACAGCGTTTCAGAGACAAG 254
Qy 339 gttaatgaagtcaactgagaccatttgcagaaggcttcttgactacacgctgcacaaaga 398
Dy 255 GCTGGAAGAGGCCTTAGAGAAATTTATGTGAGCGGATCTGGACTATAGTGTTCAGCGCTGA 314
Qy 399 gaggactggcagcaaacgggtttgccaaagggtatctcgagagacctttgagacgctgcacaa 458
Dy 315 GCGCAAGGGCTCAGTGAGATATGCCAAGGGTCAGAGTCAGACCATGGCAACACTGAAGG 374
Qy 459 cctagtccacaaagggtcgaagggtgatgatgatatccctatgagctgtggaacgagac 518
Dy 375 CCTAGTCAGAAGGGGTGAAGGTGATCTGGGGATCCCTCTGGAGCTTTGGGATGAGCC 434
Qy 519 ctacagcagaggtgctgaccccaagaagcagtgatgacgtgctggtggaagactttgaaga 578
Dy 435 CAGCGTGGAGGTCACATACCTCAAGAAGCAGTGTGAGACCATGTTGGAGGAGTTTGAAGA 494
Qy 579 ggtgattgagactggttacagaaccacaggaagacacctgactgactgaattcctctgtgc 638
Dy 495 CATGTGGGAGACTGCTACTTCCACCATCAGGAGGAGCGCCCTACAAAATTTTCTCTGTGA 554
Qy 639 caaccacgtctgaagggaaggaagcacagagttgcctagcagagcggtgctcgtggcaagaa 698
Dy 555 AGGTCATGTGCTCCAGCTGCTGAAACTGCATGTCTACAGGAAACTTGSACTGGAAGGA 614
Qy 699 g 699
Dy 615 G 615

RESULT 13

AX136616 AX136616 746 bp DNA linear PAT 30-MAY-2001
LOCUS Sequence 538 from Patent EP1067182.
DEFINITION
ACCESSION AX136616
VERSION AX136616.1 GI:14273020

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 746)

AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.

TITLE Secretory protein or membrane protein

JOURNAL Patent: EP 1067182-A 538 10-JAN-2001;

FEATURES Location/Qualifiers

1..746

source /organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 201 a 142 c 251 g 149 t 3 others

ORIGIN

Query Match 8.7%; Score 160.2; DB 6; Length 746;
Best Local Similarity 57.3%; Pred. No. 1.4e-29;
Matches 310; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

Qy 162 ggctgaggagaccgactgggtgcgattgccagcaaatgcgaagtgtgcaagtattgtgc 221
Dy 132 GGAGGAGCAGCATGACACAGAACGGCTTGCACGAGCAATGCCGAAGGTGTGTAAGCTGCTGAG 191
Qy 222 tgtgagctgaagtgcgcttttgcgaaggaagaaagacccaaggaagtgtgacaccgg 281
Dy 192 CACAGAGTACAGGCGGAAGTGTGAGTCCGACCGTGCATCTCGAGAGGTGCTCGAGCTGGG 251

Qy 282 ctatggcatcctgcacgggaagggtcctggagtcaag---tacaccaagtcggacttacg 338
Dy 252 GCAGGTGCTGGATACAGGCAAGGAGGAGACACAGTGCCTTACAGCGTTTCAGAGACAAG 311

Qy	339	g	ttaatggaagtcactgagaccattgtccaaagagccttcggactacacgctgcacaaagg	398
Db	312	G	CTGGAAGAGGCCTTAGAGAATTTATGTGAGCGGATCCTGGACTATAGTGTTCAGCTGA	371
Qy	399	g	agactggcagcaaacggtttgccaaaggtatgctcgagaccittgagacgctgcacaa	458
Db	372	G	CGCAAGGCTCACTGAGATATGCCAAGGGTCAAGTCAAGACCATGGCAACACTGAAGG	431
Qy	459	c	ctagtccacaaaggggtcaaggtggtgatggatatccctcatgagctgtggaaacgagac	518
Db	432	C	CTAGTCAGAAGGGGTGAAGTGGATCTCGGGATCCTCTGCAGCTTTGGGATGAGCC	491
Qy	519	ctc	gcagagagtggtgctgaacctcaagaagcagtgtagcgtgctgtggaagagtttgaaga	578
Db	492	C	AGGCTGGAGGCTACATACCTCAAGAAGACAGTGTGAGACCATGTTGGAGGAGTTGA	551
Qy	579	g	gtatgagactgggtacagaaaccaccagagagacacctgactgaattcctctatgc	638
Db	552	C	ATTGTGGGAGACTGGTACTTCCACCATCAGGAGCAGCCCTACAAAATTTCTCTGTGA	611
Qy	639	ca	ccacgtgctgaagggaagacagcagtgctgctagcagagcgggtgctggccaagaa	698
Db	612	A	GGTCATGTGCTCCAGCTGCTGAACATGTCATGCTACAGGAACATTGGACTGCAAGGA	671
Qy	699	g	699	
Db	672	G	672	
RESULT 14				
BD006701				
LOCUS	BD006701		747 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Novel polypeptide.			
ACCESSION	BD006701			
VERSION	BD006701.1 GI:18635072			
KEYWORDS	JP 2001029090-A/4.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	Ito,Y., Mogi,S., Tanaka,H., Okubo,S. and Ogi,K.			
TITLE	Novel polypeptide			
JOURNAL	Patent: JP 2001029090-A 4 06-FEB-2001;			
COMMENT	TAKEDA CHEMICAL INDUSTRIES LTD OS Homo sapiens (human) PN JP 2001029090-A/4 PD 06-FEB-2001 PE 19-MAY-2000 JP 2000147530 PR PI YASUAKI ITO, SHINTCHI MOGI, HIDEYUKI TANAKA, SHOICHI OKUBO, PI KAZUHIRO OGI PC C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61K48/00, PC A61P1/00, PC A61P5/00,A61P9/00,A61P11/00,A61P25/00,A61P25/18,A61P31/00, PC A61P35/00, PC A61P37/00,C07K14/47,C07K16/18,C12P21/02,G01N33/15,G01N33/50, PC G01N33/566// PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/08,(C12P21/02, PC C12R1:91), PC (C12N5/10,C12R1:91),(C12P21/08,C12R1:91),C12N15/00,A61K37/02, CC (C12N5/00,C12R1:91) CC CC			
FEATURES	Key Location/Qualifiers			
source	1..747			
FT	/organism='Homo sapiens (human)'			
FT	Location/Qualifiers			
source	1..747			
BASE COUNT	209 a 143 c 255 g 140 t			

Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 42 Row: j Column: 23.

FEATURES

Location/Qualifiers
1..1413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4991480"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
c1..774

CDS

/codon_start=1
/product="Unknown (protein for IMAGE:4991480)"
/protein_id="AAH19903.1"
/db_xref="gi:18043799"
/translation="QAVIVCRFRVMPVRLGILLFLAVHEAWAGMLKEEDDDTER
LPSKCEVCKLLSTELQAELSRTGRSREVLELQVLDTGKRRHVPYSVSETRLEEALD
NLCEIRIDYSVHAERKSLRYAKGQSQTMLKGLVQKGVKVDIGIPILEWDEPSVEV
TYLKKQCTEMLDEEDIVGDWYFHQSQPLQNFCEHVLPAETAACLOETWTGKEIT
DGEKTECEEEEEEDEEGGDKMTKTSHPKLDREDL"

BASE COUNT 378 a 301 c 413 g 321 t
ORIGIN

Query Match		8.7%	Score 160.2	DB 9	Length 1413
Best Local Similarity		57.3%	Pred. No. 1.4e-29		
Matches 310		Conservative 0	Mismatches 228	Indels 3	Gaps 1
Qy	162	ggctgagagaccgactgggtgctgattgctccagcaaaatgcaagtgatgcaagtatgtgc	221		
Db	108	CGAGGAGGACGNTGACAGAACGGTTGCCACGAATGCCGAAGTGTGTAGCTGCTGAG	167		
Qy	222	tgtgagctgaagtcggcttttgagaaacggaagcaaggaagtgattgacacccgg	281		
Db	168	CACACAGCTACAGGCGGAACAGTGCACCGTCGATCTCGAGAGGTGCTGGAGCTGGG	227		
Qy	282	ctatggatcctggaaggggctctggagtaag---tacaccaagtoggacttacg	338		
Db	228	GCAGGTGCTGGATACAGCAAGAGGAGAGACACGTGCCTTTACAGCGTTTCAGAGACAAG	287		
Qy	339	gttaattgaagtcactgagaccatttgaagggctctggaactacagcctgcacaa	398		
Db	288	CCTGGAAGAGGCTTAGAGAAATTTATGTGACGGGATCTGGACTATAGTTTCACGCTGA	347		
Qy	399	gaggactggcagcaaccgggttgccaagggtatgtcggaagcctttgagagcgtgcacaa	458		
Db	348	CGCCAAGGCTCACAGATATGCCAAGGGTCAGAGTCAGACCATGCCAACACTGAAGG	407		
Qy	459	cctagtcacaaaggggtcaagggtggtgatgatatcccttatgagctgtggaacgagac	518		
Db	408	CCTAGTCAGAGAGGGGGTGAAGGGTGATCTGGGGATCCCTCTGGAGCTTTGGGATGAGCC	467		
Qy	519	ctcagcagaggtggctgaccccaagaagcagtgctgctggtggaagatttgaaga	578		
Db	468	CAGCTGGAGGTCACATACCTCAAGAACGAGTGTGAGACCATGTTGGAGGAGTTTGAAGA	527		
Qy	579	gggtgattgagactgggtacaggaaccaccaggaaggaagcctgactgaattcctctgtgc	638		
Db	528	CATTCGGAGACTGCTACTCCACCATCAGGAGCAGCCCTACAAAATTTCTCTGTGA	587		
Qy	639	caaccacgtgctgaaggaaggaacacaggttgcctagacagagcgtgtgtgcaagaa	698		
Db	588	AGGTATGCTGCTCCAGCTGCTGAAACTGCATGTCTACAGGAAACTTGGACTGGAAGGA	647		

Qy 699 g 699
Db 648 G 648

Search completed: May 24, 2002, 00:14:24
Job time: 11370 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 23:05:57 ; Search time 4203.75 Seconds
(without alignments)
1512.237 Million cell updates/sec

Title: US-09-696-686-50
Perfect score: 471
Sequence: 1 acagaaacagaacacaaa.....atggagtcctgctgagct 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
1	295.2	62.7	558 12	BH332134	CH230-201
2	159.8	33.9	416 9	AW245217	AW245217 2819660.5
3	159.8	33.9	515 10	BE393833	BE393833 601312107
4	159.8	33.9	676 10	BE383072	BE383072 601298881
5	159.8	33.9	690 10	BI835105	BI835105 603087770
6	159.8	33.9	828 10	BE704841	BE704841 602688484
7	159.8	33.9	848 10	BE908458	BE908458 601503090
8	159.8	33.9	1602 11	BC011767	BC011767 Homo sapi
9	159.6	33.9	695 10	BI760437	BI760437 DREF2p547B
10	158.2	33.6	615 9	AL134371	AL134371
11	158.2	33.6	695 10	BI906932	BI906932 603064743
12	158.2	33.6	706 10	BI766238	BI766238 603052885
13	158.2	33.6	827 10	BG827636	BG827636 602748652
14	158.2	33.6	951 10	BG755172	BG755172 602711494
15	156.2	33.2	724 10	BI523540	BI523540 603175461
16	153.4	32.6	646 12	AG048758	AG048758 Pan trogl
17	152.8	32.4	822 10	BI770517	BI770517 603055496

18	149.4	31.7	796	10	BG116444	BG116444 602317506
19	147.8	31.4	799	10	BI114340	BI114340 602862372
20	145.8	31.0	828	9	AL524223	AL524223 AL524223
21	144.8	30.7	318	12	BH332555	BH332555 CH230-201
22	142.4	30.2	780	10	BI755826	BI755826 603024884
23	141	29.9	630	9	AL534027	AL534027 AL534027
24	128.8	27.3	405	9	AA065263	AA065263 C06502r T
25	122	25.9	405	9	AA065262	AA065262 C04502r T
26	107.8	22.9	483	9	BB857917	BB857917 BB857917
27	107	22.7	679	10	BG762655	BG762655 602734550
28	103.4	22.0	925	9	AL520942	AL520942 AL520942
29	103	21.9	362	10	R24856	R24856 Y929F01.r1
30	100.2	21.3	898	10	BE741659	BE741659 601594707
31	100	21.2	643	9	BB661603	BB661603 BB661603
32	98	20.8	599	10	BF780486	BF780486 602104020
33	94.4	20.0	472	9	BB863087	BB863087 BB863087
34	94.4	20.0	776	10	BG864846	BG864846 602799277
35	93	19.7	637	9	BB650071	BB650071 BB650071
36	93	19.7	649	9	BB652011	BB652011 BB652011
37	91	19.3	1553	11	AK005532	AK005532 Mus muscu
38	90	19.1	1393	10	BF316156	BF316156 601896158
39	90	19.1	1586	11	AK010702	AK010702 Mus muscu
40	89	18.9	623	9	BB661026	BB661026 BB661026
41	87.4	18.6	502	9	BB853356	BB853356 BB853356
42	86	18.3	821	10	BI653380	BI653380 603301117
43	85	18.0	330	9	BB856691	BB856691 BB856691
44	84.8	18.0	358	9	AW357718	AW357718 41330 MAR
45	83	17.6	838	10	BF534535	BF534535 602048047

ALIGNMENTS

RESULT 1

BH332134
LOCUS BH332134 558 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-201K14.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-201K14, DNA sequence.
ACCESSION BH332134
VERSION BH332134.1 GI:17262848
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 558)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartabeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-201K14.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 201 row: K column: 14
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..558
/organism="Rattus norvegicus"
/strain="BN/SSNHsd/MCW"

FEATURES

source

www-bio.lnlnl.gov/bbrp/image/image.html Base Calling / Quality
 Scores: PHRED from University of Washington Genome
 Center
 Trimming: cross_match from University of Washington
 PHRAP suite, Poly-T Identification: patMatch.pl from
 Dr. Philippa Genome Project, University of Washington
 Genome Center: <http://www.genome.washington.edu>
 plate: LICM2 row: C column: 21
 High quality sequence stop: 361

FEATURES	Location/Qualifiers
source	1. .416

```

I. 140
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2819660"
/clone_lib="NIH_MGC-7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"

```

```

/lab_host="DH10B phage-resistant"
/notes="lung; Vector: pORF87; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size selected >500bp for average
insert size 1.8kb. Library
constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using
ZAP-cDNA synthesis kit
from Life Technologies."

```

BASE COUNT
ORIGIN

Query Match	33.9%	Score 159.8	DB 9	Length 416
Best Local Similarity	71.6%	Prod. No. 2.8e-25		
Matches 239	Conservative	0	Mismatches 88	Indels 7
137	ggcagtgatcgtcataggcgccaaagccaccatccgctctctgattgggtgagatgg	196		
89	GGAAAGGACTTTGCCATAGTTCGTGAGGCCACCATCTGCTCTTACTTGCCCAAGGGCG	148		
197	gaaaaaaaagatagttcctctctatggtcttaagcagacgcgcgagcgaaacccattgg	256		
149	TAAAGATAG-----TCCTCCCATTAGCTAGAGAGCAACACCC-AGAAAGCCTATTGG	201		
257	ttngtgcgcgcgcgcgccttgcgtttccgaagcgcgtagagctaccgggcgaagggg	316		
202	CTGCGCGCTCGCGGGCCCTTGGTCCGCTTTGAAGCGGGCTCGCGGTGCAGAGGAGGGC	261		
317	cgggcgcgagctcgcggttgcggtttaccagagacacgtgcagtcgccgcgaagcgg	376		
262	GGCGCGGAGGCTAGCTGGTTGTCTGTGGTTGCTCGGAGGCACGTGTGCAGTCCCGGAAGCG	321		
377	cggggggaagctgctccgcgcgcgtccgcggaaggaagccgcgcgggtccgcgtctctct	436		
322	CAGAGGGAACCTCTCCGCGCGCGCCGCGGAGGAGAACCCGCCGCTCTCTTAGGGTCC	381		
437	gggtccgcgctgggccatggagtcacatgtctgagc	470		
382	GGGCGCGCGCGGCCCATGGATTCAATGCGCTGAGC	415		

3	RESULT
BE393833	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	

BE393833 515 bp mRNA linear EST 21-JUL-2000
601312107F1 NTH_MGC_44 Homo sapiens cDNA clone IMAGE:3658944 5',
mRNA sequence.
BE393833
BE393833.1 GI:9339198
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 515)
NTH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health. Mammalian Gene Collection (MGC)

[illegible]

ACCESSION	BC011767	GI:15079962
VERSION	BC011767.1	
KEYWORDS	HTC.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 1602)	
TITLE	Strausberg, R.	
JOURNAL	Direct Submission	
REMARK	Submitted (30-JUL-2001) National Institutes of Health, Mammalian	
COMMENT	Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
	NIH-MGC Project URL: http://mgc.nci.nih.gov	
	Contact: MGC help desk	
	Email: cgapbs-r@mail.nih.gov	
	Tissue Procurement: ATCC	
	cDNA Library Preparation: Rubin Laboratory	
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),	
	Calithersburg, Maryland:	
	Web site: http://www.nisc.nih.gov/	
	Contact: nisc.mgc@nih.gov	
	Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blackesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karling, E., Legaspi, R., Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McLoakey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantropop, S., Thomas, P.J., Tlionson, E.E., Touchman, J.W., Tsaurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.	
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
	Series: IRAL Plate: 27 Row: f Column: 20	
	This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein	
	This clone has the following problem: incomplete processing.	
FEATURES	Location/Qualifiers	
source	1..1602	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3609644"	
	/tissue_type="Uterus, endometrium adenocarcinoma"	
	/clone_lib="NIH_MGC_44"	
	/lab_host="DH10B-R"	
	/note="Vector: pOTB7"	
BASE COUNT	361 a 462 c 491 g 288 t	
ORIGIN		
Query Match	33.9%; Score 159.8; DB 11; Length 1602;	
Best Local Similarity	71.6%; Pred. No. 2.9e-25;	
Matches	239; Conservative 0; Mismatches 88; Indels 7; Gaps 2;	
QY	137 ggcagtgatgcgtataggcgccaaagccaccatccgctctctgattgggtgagatgg 196	
Db	11	
81	GGAAAGGACTTGGCATAGTCGTGTAGCGCCACCACCATCTGCTCTTACTGTCGAAGGGCG 140	
QY	197 gaaaaaaaaagatgtcctctctatggctataaagcagacgcgcgagcgaacccattgg 256	
Db	11	
141	TAAAGATAG- ----TCCTCCCATTAGCTAGAGACGAACCC- AGAAGACCTATTGG 193	
QY	257 ttngtgcgcgcgcggccttggctgggttgcgaagcgcgctagagctaccgagggg 316	
Db	11	
194	CTGCGCGCTCGCGGGCGCTTGGTCCGCTTTGAAGCGGGGCTCGCGCTCGAGAGAGGGCG 253	
QY	317 cgggcgcgagctcgcgttgccttgggttaccacagagacacgtgcgcagtcgccgaagcgg 376	
Db	11	
254	GGCGGGAGGCTAGCTGTGTGCTGGTTCGTGGAGGACGCTGTGTGCAATCCCGGAAGCGG 313	

Db	153	TGCTGCGCGCTCCGCGGGCCCTTGGTCCGCTTTTGAAGCGCGGCTGCGGCTGCGAGAGAG	212			
Qy	314	ggcgggcgagctccqcttgctgctgattaccagagacacgtgqgcagctcccggaag	373			
Db	213	GGCGGGGAGGCTAGCTGTTGCTGCTCGGAGGACAGTGTGCAGTCCCGGAG	272			
Qy	374	cggcgggggaagctctccgcgcgcgtgcccggaggagcgcgcgggtccgcctctgc	433			
Db	273	CGCGAGGGGAAACTCTCCGCGCGCGCGCGGAGGAGAAACCGCGCTCTTTAGG	332			
Qy	434	tcgggtcggctggccatgagctcctgctgctgagc	470			
Db	333	TCCGGCCCGCGCGGCCCATGGATTCAATGCGCTGAGC	369			
RESULT 10						
AL134371						
LOCUS	AL134371	615 bp	mrna linear EST 25-FEB-'97			
DEFINITION	DKFp547B155_r1 547 (synonym: hfb1) Homo sapiens cDNA clone					
ACCESSION	AL134371					
VERSION	AL134371.1					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 615)					
TITLE	Blum,H., Bauersachs,S., Mewes,W., Weil,B. and Wiemann,S.					
JOURNAL	EST (Blum,H., Bauersachs,S., Mewes,H.W., Weil,B. and Wiemann,S.					
COMMENT	Unpublished (1999) Contact: Blum H MPS Am Klopferspitz 18a D-82152 Martinsried, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by LMU (Ludwig Maximilians University, Munich/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available. This clone (DKFp547B155) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.					
FEATURES						
source						
1..615						
/organism="Homo sapiens"						
/db_xref="taxon:9606"						
/clone="DKFp547B155"						
/clone_lib="547 (synonym: hfb1)"						
/tissue_type="brain"						
/dev_stage="fetal"						
/lab_host="xl-2blue"						
/note="Vector: pMP1; Site_1: NotI; Site_2: SalI"						
BASE COUNT	103 a	178 c	211 g			
ORIGIN	123 t					
Query Match 33.6%; Score 158.2; DB 9; Length 615;						
Best Local Similarity 71.3%; Pred. No. 6.3e-25;						
Matches 238; Conservative 0; Mismatches 89; Indels 7; Gaps						
Qy	137	ggcagtgatcgctataggcgccaaagcacaccatccgctctctgattggtagatgg	196			
Db	77	GGAAGGACTTTGCCATAGTGCCTGAGCGCCACCATTGCTCTTACTGGCCAAGGGCG	136			
Qy	197	gaaaaaaaagatgattctctctattggtctataagcagacgcgcgacccattgg	256			
Db	137	TAAAGATAG-----TCTTCCCATAGCTAGAGACAAAC-CCCAGAAAGCCTATTGG	189			
Qy	257	ttngtgcgcccgggcctttggttcggttcgaagccgcgtagaggtaccggcgagggg	316			
Db	190	CTGCGCGCTCCGCGGCTTGTGCTCGCTTTCAAGCGCGGCTGCGCTGCGAGAGAGGC	249			

RESULT 12

BI766238	706 bp	mrna	linear	EST 25-SEP-2005			
603052895F1	NIH_MGC_122	Homo sapiens	CDNA clone	IMAGE:5202525			
mrna sequence.							
BI766238							
BI766238.1	GI:15757816						
EST.							
human.							
Homo sapiens							
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
1 (bases 1 to 706)							
NIH-MGC http://mgc.ncl.nih.gov/ .							
National Institutes of Health, Mammalian Gene Collection (MGC)							
Unpublished (1999)							
Contact: Robert Strausberg, Ph.D.							
Email: cgabbs@remail.nih.gov							
Tissue Procurement: Life Technologies, Inc.							
CDNA Library preparation: Life Technologies, Inc.							
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)							
DNA Sequencing by: Incyte Genomics, Inc.							
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov							
row: 1 column: 22							
High quality sequence stop: 706.							

```

BASE COUNT      130 a      203 c      233 g      139 t      1 others
ORIGIN
FEATURES
SOURCE
    1. .706
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:5202525"
    /clone_lib="NIH_MGC_122"
    /lab_host="DH10B"
    /note="Organ: pooled lung and spleen; Vector: pCMV-SP"
    Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
    anonymous pool of 24 week female lung, 16 week female
    spleen, and 20-22 week male spleens. Library is oligo
    primed and directionally cloned (EcoRV site is destroyed
    upon cloning). Average insert size 1.4 kb, insert size
    range 1-3 kb. Library is normalized and enriched for
    full-length clones and was constructed by C. Gruber
    (Invitrogen). Research Genetics tracking code 026.
    this is a NIH_MGC Library."

```

BASE COUNT 130 a 203 c 233 g 139 t 1 others

ORIGIN

Query Match 33.6%; Score 158.2; DB 10; Length 706;
Best Local Similarity 71.3%; Pred. No. 6.4e-25;
Matches 238; Conservative 0; Mismatches 89; Indels 7; Gaps

Qy 137 ggcagtgcatacgtataggcgccaaagccaccatccgctctctctgattgggtgcagatgg 1961
Db 76 GGAAGGACATTTGGCATAGGTCGGCTGAGGCGACCATCTGCTCTTACTGCCNAGGGCG 1351

Qy	137	ggcagtgatatacgctataggcgccaagccaccatccgctctctgattgggtgagatgg	196
Db	67	ggaagagactttggcattagcttgagggccaccatctgcttcttctgagcccaaggcgc	126
Qy	197	gaaaaaaaagatagttcctctcatggctataaagcagacgcgagcgaaacccattgg	256
Db	127	taaaaagatag-----tccctccattatgctagagagcaaacccc-agaagccctattgg	179
Qy	257	ttngtgcgcgcgggcttggcttgcgttgcgaagccgctagagctaccggtgcgagggg	316
Db	180	ctgcccctgcggcgcccttggtccgctttgaagcggcgctgcgcgcgcgagagggc	239
Qy	317	cggcccgagactcgccttgcgttggttaccagagacacgctgcgagtcgcggaagcgg	376
Db	240	ggcgggagagctactcgttgcgttggttcgagagcagctgctgcaatcccggaagcgg	299
Qy	377	ccggggaaagctgctccgcgcgcgtccgagagaaagccgcgcgcgtcgcctctgctct	436
Db	300	cgagggaactgcttcgagcgcgcggcgagagaaaccccggtctctttaggtgcc	359
Qy	437	gggtccggctggggccatggagtcctctctgagc	470
Db	360	ggcggcgccggcgcatggattcaatgcctgagc	393

RESULT 14

BG755172

LOCUS

DEFINITION

602711494F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851982 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 951)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabp-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLCMI695 row: k column: 23

High quality sequence stop: 831.

Location/Qualifiers

1..951

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4851982"

/clone_lib="NIH_MGC_48"

/tissue_type="primary B-cells from tonsils (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

193 a 258 c 300 g 200 t

BASE COUNT

ORIGIN


```

US-09-423-890-7
; Sequence 7, Application US/09423890
; Patent No. 6312934
; GENERAL INFORMATION:
; APPLICANT: CADUS PHARMACEUTICAL CORPORATION
; TITLE OF INVENTION: HUMAN MERK PROTEIN AND NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: CPI-085CPPC
; CURRENT APPLICATION NUMBER: US/09/423,890
; CURRENT FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: USSN 60/078,153
; PRIOR FILING DATE: 1998-03-16
; PRIOR APPLICATION NUMBER: USSN 60/099,165
; PRIOR FILING DATE: 1998-09-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 5253
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(493)
US-09-423-890-7

Query Match      8.2%; Score 38.4; DB 4; Length 5253;
Best Local Similarity 56.2%; Pred. No. 0.075;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 292 ccgctagaggtctacggcgagggcgagggcgagagctgcggtgcggtaccagga 351
      |||| |||| | |||| |||| | |||| |||| | |||| |||| |
Db 3 ccgcgagagaaatgcgcgcgcgagcgatcgcgctctgcgggattcccggc 62

QY 352 gacagctgcgagtcgccggaagcgagggggaagctgctccgcgcgtcccgagga 411
      |||| |||| | |||| |||| | |||| |||| | |||| |||| |
Db 63 gccgcggcgcgagtcgccgagggcgagggcgagggcgagaggagagctctccagga 122

QY 412 agcgcgc 419
      |||| |||
Db 123 agcgcgc 130

RESULT 3
US-08-141-893-1
; Sequence 1, Application US/08141893
; Patent No. 5489519
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/141,893
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923; 08/029,340
; FILING DATE: 27-OCT-1992; 8-MAR-1993
; ATTORNEY/AGENT INFORMATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/407,207
; FILING DATE: 20-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steeg, Carol Miernicki
; REGISTRATION NUMBER: 39,539
; REFERENCE/DOCKET NUMBER: Q1546
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2342
; TELEFAX: (613) 545-6853
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..4788
; US-08-463-092B-1

```

```

Query Match      7.8%; Score 36.8; DB 1; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 309 gcaagggcgccggagctccggtggttaccagagacagtcgcagtgccc 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 114 GCCCGGTGCCCGCGCCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 173

Qy 369 gaaagcgccgggggaagctgctccgctgcgcgaggaagcgccggtccgc 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 CGCCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 233

Qy 429 tctgctctggtccgg 444
    ||||| ||||| |||||
Db 234 CCCGCTCTGGGACTGG 249

```

```

RESULT 5
US-08-463-092B-3
; Sequence 3, Application US/08463092B
; Patent No. 5766880
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,092B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/029,340

```

```

; FILING DATE: 8-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/141,893
; FILING DATE: 26-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/407,207
; FILING DATE: 20-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steeg, Carol Miernicki
; REGISTRATION NUMBER: 39,539
; REFERENCE/DOCKET NUMBER: Q1546
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2342
; TELEFAX: (613) 545-6853
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..4788
; US-08-463-092B-3

Query Match      7.8%; Score 36.8; DB 1; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 309 gcaagggcgccggagctccggtggttaccagagacagtcgcagtgccc 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 114 GCCCGGTGCCCGCGCCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 173

Qy 369 gaaagcgccgggggaagctgctccgctgcgcgaggaagcgccggtccgc 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 CGCCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 233

Qy 429 tctgctctggtccgg 444
    ||||| ||||| |||||
Db 234 CCCGCTCTGGGACTGG 249

RESULT 6
US-08-462-109A-1
; Sequence 1, Application US/08462109A
; Patent No. 5882875
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: MULTIDRUG RESISTANT TUMOR CELLS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,109A
; FILING DATE:
; CLASSIFICATION: 435

```

7
US-08-462-109A-3
; Sequence 3, Application US/08462109A
; Patent No. 5882875
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: MULTIDRUG RESISTANT TUMOR CELLS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,109A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

RESULT      8
US-08-460-907B-1
; Sequence 1, Application US/08460907B
; Patent No. 5891724
; GENERAL INFORMATION:
; APPLICANT: Deeley, Roger G.
; APPLICANT: Cole, Susan P.C.
; TITLE OF INVENTION: METHODS FOR CONFERRING MULTIDRUG
; TITLE OF INVENTION: RESISTANCE ON A CELL
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; City: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,907B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923

```



```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,179A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: PQI-002CP8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-463-179A-3

Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 309 gcgaggggcgccggagctgcgcgttgcctggttaccacagagacacgtgcgcagtcgcc 368
Db 114 GCCCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173
QY 369 gaaagcgccgggggaagctctccgcgcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 174 CGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233
QY 429 tctgctctgggtccgg 444
Db 234 CCCGCTCTGGGACTGG 249

RESULT 12
US-08-461-384B-1
; Sequence 1, Application US/08461384B
; Patent No. 6025473
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,179A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: PQI-002CP8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-463-179A-1

Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 309 gcgaggggcgccggagctgcgcgttgcctggttaccacagagacacgtgcgcagtcgcc 368
Db 114 GCCCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173
QY 369 gaaagcgccgggggaagctctccgcgcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 174 CGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233
QY 429 tctgctctgggtccgg 444
Db 234 CCCGCTCTGGGACTGG 249

RESULT 11
US-08-463-179A-3
; Sequence 3, Application US/08463179A
; Patent No. 6001563
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING CHEMOSENSITIZERS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;; SOFTWARE: ASCII text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/461,384B
;; FILING DATE: 05-JUN-95
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/966,923
;; FILING DATE: 27-OCT-1992
;; APPLICATION NUMBER: 08/029,340
;; FILING DATE: 8-MAR-1993
;; APPLICATION NUMBER: 08/141,893
;; FILING DATE: 26-OCT-1993
;; APPLICATION NUMBER: 08/407,207
;; FILING DATE: 20-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Steeg, Carol Mlernicki
;; REGISTRATION NUMBER: 39,539
;; REFERENCE/DOCKET NUMBER: Q1547
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (613) 545-2342
;; TELEFAX: (613) 545-6853
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5011 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 196..4788
US-08-461-384B-1
```

```
Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 309 gcaaggggcgagcgagcgttccgtggttaccagagacagtcgcagtcgc 368
Db 114 GCCCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173

Qy 369 ggaagcgccgggggaaagctctccgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 174 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233

Qy 429 tctgctctgggtcccg 444
Db 234 CCCGCTCTGGGACTGG 249
```

```
RESULT 13
US-08-461-384B-3
; Sequence 3, Application US/08461384B
; Patent No. 6025473
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,384B
```

```
;; FILING DATE: 05-JUN-95
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/966,923
;; FILING DATE: 27-OCT-1992
;; APPLICATION NUMBER: 08/029,340
;; FILING DATE: 8-MAR-1993
;; APPLICATION NUMBER: 08/141,893
;; FILING DATE: 26-OCT-1993
;; APPLICATION NUMBER: 08/407,207
;; FILING DATE: 20-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Steeg, Carol Mlernicki
;; REGISTRATION NUMBER: 39,539
;; REFERENCE/DOCKET NUMBER: Q1547
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (613) 545-2342
;; TELEFAX: (613) 545-6853
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5011 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 196..4788
US-08-461-384B-3
```

```
Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 309 gcaaggggcgagcgagcgttccgtggttaccagagacagtcgcagtcgc 368
Db 114 GCCCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 173

Qy 369 ggaagcgccgggggaaagctctccgcgcgcgcgcgcgcgcgcgcgcgc 428
Db 174 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 233

Qy 429 tctgctctgggtcccg 444
Db 234 CCCGCTCTGGGACTGG 249
```

```
RESULT 14
US-08-407-207A-1
; Sequence 1, Application US/08407207A
; Patent No. 6063621
; GENERAL INFORMATION:
; APPLICANT: Deeley, Roger G.
; APPLICANT: Cole, Susan P.C.
; TITLE OF INVENTION: ANTIBODIES TO A MULTIDRUG RESISTANCE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/407,207A
; FILING DATE: 20-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: Q1512
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-407-207A-1

Query Match 7.8%; Score 36.8; DB 3; Length 5011;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 309 gcaggggcccgcgagctccgttgcgtggttaccagagacacgtgcgagtcgcc 368
Db 114 GCCCGTGCCCGCCGCCCGCCGAGCAACCGGCCCGATCACCCGCCCGCGTGCC 173
QY 369 ggaagcgcgggggaagtgtctgcgcgcgtgcgaggaagcgcgcgggtccgc 428
Db 174 CGCCGCCGCCCGCCACCGCATGCGCTCCGGGCTTCGACGCGCGATGGCTCCGA 233
QY 429 tctgtctgggtccgg 444
Db 234 CCCGCTCTGGGACTGG 249

RESULT 15

US-08-586-165-1
Sequence 1, Application US/08586165
Patent No. 6054298

GENERAL INFORMATION:

APPLICANT: Laufer, Edward M.
APPLICANT: Orozco, Olivia E.
APPLICANT: Tabin, Clifford J.
TITLE OF INVENTION: Fringe Proteins and Pattern Formation
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/586.165
FILING DATE: 16-JAN-1996
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: HU95-05

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: Join(1..240, 244..474, 478..531, 538..579, 583
LOCATION: ..678, 682..687, 691..807, 811..843, 850..870,
LOCATION: 874..990, 994..1056, 1060..1083, 1087..1104, 1108
LOCATION: ..1119)
US-08-586-165-1

Query Match 7.6%; Score 36; DB 3; Length 1120;
Best Local Similarity 50.6%; Pred. No. 0.2;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 284 ttcgcaagccgctagaggctaccggcgagggcgggcgagagctgcggttgcggtgt 343
Db 65 TTCTCTGCTGCTGCTGCCCGGGGACAGCCCCCGCGCGCGCGCGCGCGCGCG 124
QY 344 taccagagacacgtgcagtcctccggaagcgcgcggggaagctgctccgcgcgctg 403
Db 125 CCCCGCGGCCCGCAGCAGGCCCTCCCGAAGCGGAGCGCGCGCGCGCGCGCGCG 184
QY 404 ccggaggaagcgcgcgggttcctctctctctggttcggttcggttcggttcggttcg 455
Db 185 CCCGGGACCCGCGCGCGCTCGGAGCGCGGGGGCGCGCGCGCGCGCGCGCGCG 236

Search completed: May 24, 2002, 00:20:45
Job time: 11711 sec

Result No.	Score	% Match	Length	DB	ID	Description
1	470	99.8	471	22	AA505284	T243 gene homolog
2	158.2	33.6	1675	20	AA521311	Human CDNA sequenc
3	148	31.4	1657	20	AA504377	Human secreted pro
4	66	14.0	1839	22	AA505278	Murine trinucleoti
5	66	14.0	1848	22	AA505286	DNA sequence of an
6	46.8	9.9	2829	21	AA575235	Human ORF790
C 7	39.6	8.4	48037	22	AA584729	Human immune/hae
C 8	39.6	8.4	48037	22	AA585963	Human immune/hae
C 9	39.6	8.4	48045	22	AA584730	Human immune/hae

Novel transgenic animals useful as animal model for characterization of function of a gene encoding trinucleotide repeat proteins (TRPs).

PT contains heterozygous disruption in a gene encoding TRP -

XX Claim 11; Fig 14; 106pp; English.

XX The present sequence for T243 homologous sequence #1 is generated by PCR

CC and is homologous to the T243 gene which encodes for a trinucleotide

CC repeat protein (TRP). The invention describes methods of producing

CC embryonic stem (ES) cells comprising a heterozygous disruption in a

CC target DNA sequence (preferably gene T243) encoding a TRP and of

CC producing a knockout mouse comprising a homozygous disruption in a gene

CC encoding TRP. The invention also relates to identifying agents capable of

CC affecting a phenotype of a knockout mouse. Also described are methods of

CC determining whether expansion of the trinucleotide repeat in a gene

CC encoding TRP produces a phenotypic change. The transgenic animals and the

CC cells are useful for identifying compounds capable of ameliorating

CC disease symptoms, and as test substrates for the identification of drugs,

CC pharmaceuticals, therapies and interventions which may be effective in

CC treating trinucleotide repeat disorders e.g. fragile X syndrome and

CC Huntington's disease. The animal models for trinucleotide repeat

CC disorders are ideal model systems to study the progression of disease in

CC vivo, the molecular basis of these diseases and show the features

CC observed in human disease. Using the mice, it is possible to model both

CC the pathogenic mechanism and the trinucleotide repeat instability in the

CC mouse.

XX Sequence 471 BP; 105 A; 129 C; 148 G; 88 T; 1 other;

SQ

Query Match 99.8%; Score 470; DB 22; Length 471;

Best Local Similarity 100.0%; Pred. No. 7e-125;

Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acagaaacacgaagaaacacacatgaagatagctgttctatccagggttagaatacgcga 60

Db 1 acagaaacacgaagaaacacacatgaagatagctgttctatccagggttagaatacgcga 60

Qy 61 aggtgtgttccaaagttagatgaaggttcacccgctaggaactgactccagctac 120

Db 61 aggtgtgttccaaagttagatgaaggttcacccgctaggaactgactccagctac 120

Qy 121 tgagctctcttagctgagctgatatcgctatagggcgccaaagccaccatcgctctc 180

Db 121 tgagctctcttagctgagctgatatcgctatagggcgccaaagccaccatcgctctc 180

Qy 181 tgattgggtgagtggaagaaacacacacacacacacacacacacacacacacacacac 240

Db 181 tgattgggtgagtggaagaaacacacacacacacacacacacacacacacacacacac 240

Qy 241 cgagcgacacacattggttgnctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300

Db 241 cgagcgacacacattggttgnctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300

Qy 301 gctaccggcgagggcgccgagctgcgcgttgcgcgttgcgcgttgcgcgttgcgcgttgcgcgt 360

Db 301 gctaccggcgagggcgccgagctgcgcgttgcgcgttgcgcgttgcgcgttgcgcgttgcgcgt 360

Qy 361 cgagctccggaagcgccgaggggaagctgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420

Db 361 cgagctccggaagcgccgaggggaagctgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420

Qy 421 gggctccgctctgctctggttccgctggtggccatggagtcacatgctgagct 471

Db 421 gggctccgctctgctctggttccgctggtggccatggagtcacatgctgagct 471

RESULT 2

AAS21311

ID AAS21311 standard; cDNA; 1675 BP.

XX

AC AAS21311;

XX

DT 24-OCT-2001 (first entry)

XX Human cDNA sequence encoding for PRO4409 polypeptide.

DE

XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;

KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;

KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;

KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX

OS Homo sapiens.

XX WO200140466-A2.

PN

XX 07-JUN-2001.

PD

XX 01-DEC-2000; 2000WO-US32678.

XX 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.

PR 06-JAN-2000; 2000WO-US00277.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US04914.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 20-MAR-2000; 2000WO-US07377.

PR 21-MAR-2000; 2000WO-US07532.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR 10-NOV-2000; 2000WO-US30873.

XX (GETH) GENENTECH INC.

PA

XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;

PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;

PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WL, Zhang Z;

XX WPI: 2001-408281/43.

DR P-PSDB; AAU12239.

DR

XX Isolated, secretory and transmembrane PRO polypeptide used to detect

PT other PRO polypeptides, link bioactive molecules to cells expressing

PT PRO polypeptides, and detect the presence of mammalian tumours e.g.

PT lung, breast, prostate, cervical.

PT

XX Claim 3; Fig 135; 813pp; English.

XX AAS21244-AAS21518 encode for novel human secretory and transmembrane

CC PRO polypeptides. The PRO polypeptides are useful to detect other

CC PRO polypeptides, to link bioactive molecules to cells expressing

CC PRO polypeptides, to modulate biological activities of cells expressing

CC PRO polypeptides, and to detect the presence of mammalian lung, colon,

CC breast, prostate, rectal, cervical or liver tumours by comparing PRO

CC polypeptide expression in a cell sample to that in a control sample.

CC Some of the 275 sequences are also useful to stimulate the release of

CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the

CC proliferation or differentiation of chondrocytes, the proliferation or

CC gene expression in pericyte cells, the release of proteoglycans from

CC cartilage, the proliferation of inner ear utricular supporting cells or

CC of T-lymphocytes, the release of a cytokine from peripheral blood

PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 18-AUG-2000; 2000US-0225759.
 PR 22-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226581.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234977.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246600.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX Disclosure; SEQ ID NO 39541; 3071pp + Sequence Listing; English.
 PS
 XX

PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225113.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPT; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 39542; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK64703

08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis.
Disclosure: SEQ ID NO 40796; 3071pp + Sequence Listing; English.
AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patient's own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/hematopoietic-related diseases, especially
cancers and cancer metastases of hematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/hematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAK82169
represent sequences used in the exemplification of the present invention.
Sequence 48045 BP; 12889 A; 9812 C; 10461 G; 14883 T; 0 other;

Query Match 8.4%; Score 39.6; DB 22; Length 48045;
Best Local Similarity 50.5%; Pred. No. 0.61;
Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 265 cccgcggccttgctggttttcgcaagcgcctagaggtaccggcgagggcgccgg 324
DB 420 CGCTGGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 361
QY 325 agctcgccgttgccgtgttaccagagacacgtgcgcagtcctccggaagcgccgggga 384
DB 360 GGCTTCCATGTCCGCCTGGTCCACAGCTCAGGGCCCGAGTCGCTGCTGCTGCCGCT 301
QY 385 agctgctccgcgcgcgtgcgcggaggaagcgcgcggggtccgtctctgtctggtccgg 444
DB 300 GCTGCTCCCGG 241
QY 445 ctggggccatg 454
DB 240 GGGGGCGCGTG 231
RESULT 11
AAC99044/C
ID AAC99044 standard; cDNA; 527 BP.
XX AAC99044;
AC AAC99044;
DT 09-MAR-2001 (first entry)
XX Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:272.
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
detection; diagnosis; identification; cytostatic; neuroprotective;
neurotic; immunomodulatory; relaxant; contraceptive; gynaecological;
antiinflammatory; cardiant; gene therapy; chromosome mapping;
linkage analysis; tissue identification; tissue typing; forensic;
neural; immune system; muscular; reproductive; gastrointestinal;
pulmonary; cardiovascular; renal; proliferative; ss.
OS Homo sapiens.
XX WO200055320-A1.
PN 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05989.
PF 12-MAR-1999; 99US-0124270.
PR (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
PI WPI; 2000-579444/54.
XX P-PSDB; AAB54279.
DR New nucleic acid that is a pancreatic cancer antigen for preventing,
treating, or ameliorating a medical condition, particular pancreatic
cancer, or for use in assays for diagnosing a pathological condition -
XX Claim 1; Page 708; 1379pp; English.
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
proteins, called pancreatic cancer antigens, given in AAB54008 to
AAB54466. The human pancreatic cancer antigens have cytostatic,
neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
gynaecological, cardiant and antiinflammatory activities, and can be used
in gene therapy. The polynucleotide and proteins can be used for
preventing, treating, or ameliorating a medical condition or in assays
for diagnosing a pathological condition or a susceptibility to one in a
subject. Binding partners to the proteins and the activity of the
proteins can be identified. The pancreatic cancer antigens can be used to
detect, treat or prevent pancreatic disorders, especially cancer.

us-09-696-686-50.rng

QY	431	tg	432
Db	176	TG	175

Search completed: May 24, 2002, 00:24:32
Job time: 9089 sec

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #20510.

XX

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

KW

XX Homo sapiens.

OS

XX WO200175067-A2.

XX

XX 11-OCT-2001.

PD

XX 30-MAR-2001; 2001WO-US08631.

XX

XX 31-MAR-2000; 2000US-0540217.

PR

XX 23-AUG-2000; 2000US-0649167.

PR

XX (HYSE-) HYSEQ INC.

PA

XX Drmanac RT, Liu C, Tang YT;

XX

XX WPI: 2001-639362/73.

DR

XX P-PSDB; ABG20519.

XX

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX

XX Claim 1; SEQ ID No 20510; 103pp; English.

PS

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (I) and its binding partners are useful for treating

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS94197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

XX sequence 3623 BP; 909 A; 865 C; 851 G; 993 T; 5 other;

XX

	Query Match	8.2%;	Score 38.6;	DB 23;	Length 3623;
	Best Local Similarity	50.5%;	Pred. No. 0.49;		
Matches	92; Conservative	0; Mismatches	90; Indels	0; Gaps	0;
QY	251 cattggttgntgcgcccgccgcttgcgttcgcgaagccgctaaggagtaccgggc	310			
Db	356 CTTTAGTGGTGTCATGCCCGCAGACATTGCTGCTGGTGCCTGTGAGGACGACACCAC	297			
QY	311 gaagggcggcgagactgcgcgttgcgcgttgtaccacagagacacgtgcgcagtcctcgg	370			
Db	296 GTTGGCCCTTCGGGTTCCTCCGACGACACCCGGAAGACCCCAGGACGCCGACGACCAG	237			
QY	371 aaagcgcgggggaaagctgctccgcgcgcgtgcgcggaggaaagccgcggggtccgcgtc	430			
Db	236 cttctmcaaacatcctcctccctccacacgggacagtgccgggggctgtccccccgagcgagccgacc	177			

JOURNAL

Submitted (20-JUL-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
On Apr 28, 2001 this sequence version replaced gi:11276104.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site:
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mouse.html
Contact: gnktm@capecod.bwh.harvard.edu

COMMENT

-----Summary Statistics
Center project name: ADY
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 79355 at least Q20
*Consensus quality: 75365 at least Q30
*Consensus quality: 69022 at least Q40
**Estimated insert size: agarose-PP - N/A
Quality coverage: agarose-PP - N/A
Quality coverage: 5.1 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 7962: contig of 7962 bp in length
* 7963 7982: gap of unknown length
* 7983 13924: contig of 5942 bp in length
* 13925 13944: gap of unknown length
* 13945 18990: contig of 5046 bp in length
* 18991 19010: gap of unknown length
* 19011 22501: contig of 3491 bp in length
* 22502 22521: gap of unknown length
* 22522 27077: contig of 4556 bp in length
* 27078 27097: gap of unknown length
* 27098 28819: contig of 1722 bp in length
* 28820 28839: gap of unknown length
* 28840 31224: contig of 2385 bp in length
* 31225 31244: gap of unknown length
* 31245 35093: contig of 3849 bp in length
* 35094 35113: gap of unknown length
* 35114 35114: contig of 3498 bp in length
* 38612 38631: gap of unknown length
* 38632 42281: contig of 3630 bp in length
* 42282 42282: gap of unknown length
* 44130 44130: contig of 1849 bp in length
* 44131 44150: gap of unknown length
* 44151 47059: contig of 2909 bp in length
* 47060 47079: gap of unknown length
* 47080 49496: contig of 2417 bp in length
* 49497 49516: gap of unknown length
* 49517 52236: contig of 2720 bp in length
* 52237 52256: gap of unknown length
* 52257 54270: contig of 2014 bp in length
* 54271 54290: gap of unknown length
* 54291 56271: contig of 1981 bp in length
* 56272 56291: gap of unknown length
* 56292 58849: contig of 2558 bp in length
* 58850 58869: gap of unknown length
* 58870 60621: contig of 1752 bp in length
* 60622 60641: gap of unknown length
* 60642 62255: contig of 1614 bp in length
* 62256 62275: gap of unknown length
* 63085 63105: contig of 810 bp in length
* 63086 63106: gap of unknown length
* 63106 65440: contig of 2335 bp in length
* 65441 65460: gap of unknown length

FEATURES

source

65461 67629: contig of 2169 bp in length
* 67630 67649: gap of unknown length
* 67650 69178: contig of 1529 bp in length
* 69179 69198: gap of unknown length
* 69199 70095: contig of 897 bp in length
* 70096 70115: gap of unknown length
* 70116 71821: contig of 1706 bp in length
* 71822 71841: gap of unknown length
* 71842 73506: contig of 1665 bp in length
* 73507 73526: gap of unknown length
* 73527 74749: contig of 1223 bp in length
* 74750 74769: gap of unknown length
* 74770 75963: contig of 1194 bp in length
* 75964 75983: gap of unknown length
* 75984 77905: contig of 1922 bp in length
* 77906 77925: gap of unknown length
* 77926 79236: contig of 1311 bp in length
* 79237 79256: gap of unknown length
* 79257 81332: contig of 2076 bp in length
* 81333 81352: gap of unknown length
* 81353 83169: contig of 1817 bp in length
* 83170 83189: gap of unknown length
* 83190 84096: contig of 907 bp in length
* 84097 84116: gap of unknown length
* 84117 85749: contig of 1633 bp in length
* 85750 85769: gap of unknown length
* 85770 86770: contig of 1001 bp in length
* 86771 86790: gap of unknown length
* 86791 88207: contig of 1417 bp in length
* 88208 88227: gap of unknown length
* 88228 88976: contig of 749 bp in length
* 88977 88996: gap of unknown length
* 88997 89855: contig of 859 bp in length.
Location/Qualifiers
1..89855
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-76I16"
/sex="male"
1..7962
/note="assembly_name:Contig145"
7983..13924
/note="assembly_name:Contig144"
13945..18990
/note="assembly_name:Contig143"
19011..22501
/note="assembly_name:Contig142"
22522..27077
/note="assembly_name:Contig141"
27098..28819
/note="assembly_name:Contig140"
28840..31224
/note="assembly_name:Contig139"
31245..35093
/note="assembly_name:Contig138"
35114..38611
/note="assembly_name:Contig137"
38632..42261
/note="assembly_name:Contig136"
42282..44130
/note="assembly_name:Contig135"
44151..47059
/note="assembly_name:Contig134"
47080..49496
/note="assembly_name:Contig133"
49517..52236
/note="assembly_name:Contig132"
52257..54270
/note="assembly_name:Contig131"
54291..56271
/note="assembly_name:Contig130"

Db	140	TAAAGATAG-----TCCTCCCATTTAGCTAGAGACAAACCC-AGAAAGCCTATTGG	192
Qy	257	ttgngtcgccccggggttggttcggttcgcaagcgctagagcgctaccggtcgagggg	316
Db	193	CTGCGCGCTCCGGCGCTTGTCTCGCTTTTAAAGCGCGGCTGCGCGTGCAGAGAGGGC	252
Qy	317	cggtcgagagctcgccgttgcgtgtttaccagagacacagtcgacgtcccggaagcgg	376
Db	253	GGCGGAGGCTAGCTGTGTCTGCTCGGAGGACGCTGTGTCAGTCCCGAAGCGG	312
Qy	377	ccgggggaagctgctcgcgcgctgcgcgaggaagcgccggtgcgctctctct	436
Db	313	CGAGGGGAAACTCTCCGCGCGCGCGGGAGGAGGAGCCCGCTCTTAGGGTCC	372
Qy	437	gggtcggtggcgatgagtcagtcctgctgacg	470
Db	373	GGGCGCGCGCGGCGCATGATTCATGCTGAGC	406
RESULT	3		
LOCUS	HS475N16/c	113109 bp	DNA linear PRI 10-FEB-2000
DEFINITION	Human DNA sequence from clone RP3-475N16 on chromosome 6p12.3-21.2. Contains the genes for CTGA4, pre-T cell receptor alpha, a novel protein similar to RPL7A (60S ribosomal protein L7A) and the 3' end of gene KIAA0240. Contains ESTs, STSS, GSSs and four putative CpG islands, complete sequence.		
ACCESSION	AL035587	5	GI:6002306
VERSION	AL035587.5		
KEYWORDS	HTG; CpG island; CTGA4; KIAA0240; RPL7A.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 113109)		
TITLE	Williams, S.		
JOURNAL	Direct Submission		
COMMENT	Submitted (02-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk On Sep 30, 1999 this sequence version replaced gi:5921383. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/MGP/Chr6 RP3-475N16 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: pcypac2 This sequence is the entire insert of clone RP3-475N16.		
FEATURES	Location/Qualifiers		
source	1..113109 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /map="p12.3-21.2"		

FEATURES

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

RESULT 11
AE316597

[illegible]

```

web site: http://www.jgl.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-23_41F9
-----
Summary Statistics
Consensus quality: 169456 bases at least Q40
Consensus quality: 196444 bases at least Q30
Consensus quality: 205812 bases at least Q20
Estimated insert size: 211300; agarose-fp estimation
Estimated insert size: 226654; sum-of-contigs estimation
Quality coverage: 3.43 in Q20 bases; agarose-fp estimation
Quality coverage: 3.19 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```

1	1422:	contig	Of	1422	bp	in	length
1423	1522:	gap	of	unknown	length		
1523	3563:	contig	of	2041	bp	in	length
3564	3663:	gap	of	unknown	length		
3664	4879:	contig	of	1216	bp	in	length
4880	4979:	gap	of	unknown	length		
4980	6061:	contig	of	1082	bp	in	length
6062	6163:	gap	of	unknown	length		
6162	7308:	contig	of	1147	bp	in	length
7309	7408:	gap	of	unknown	length		
7409	8607:	contig	of	1199	bp	in	length
8608	8707:	gap	of	unknown	length		
8708	10851:	contig	of	2144	bp	in	length
10852	10951:	gap	of	unknown	length		
10952	12535:	contig	of	1584	bp	in	length
12536	12635:	gap	of	unknown	length		
12637	13837:	contig	of	1202	bp	in	length
13838	13937:	gap	of	unknown	length		
13938	16329:	contig	of	2392	bp	in	length
16330	16423:	gap	of	unknown	length		
16430	18243:	contig	of	1813	bp	in	length
18243	18343:	gap	of	unknown	length		
18343	21190:	contig	of	2748	bp	in	length
21091	21191:	gap	of	unknown	length		
21191	23891:	contig	of	2701	bp	in	length
23892	23991:	gap	of	unknown	length		
23992	26366:	contig	of	2375	bp	in	length
26367	26466:	gap	of	unknown	length		
26467	29744:	contig	of	3278	bp	in	length
29745	29844:	gap	of	unknown	length		
29845	33746:	contig	of	3902	bp	in	length
33747	33846:	gap	of	unknown	length		
33847	38047:	contig	of	4201	bp	in	length
38048	38147:	gap	of	unknown	length		
38148	41062:	contig	of	2915	bp	in	length
41063	41162:	gap	of	unknown	length		
41163	44603:	contig	of	3441	bp	in	length
44604	44703:	gap	of	unknown	length		
44704	47610:	contig	of	2907	bp	in	length
47611	47710:	gap	of	unknown	length		
47711	51750:	contig	of	4040	bp	in	length
51751	51850:	gap	of	unknown	length		
51851	56139:	contig	of	4289	bp	in	length
56140	56239:	gap	of	unknown	length		
56240	63693:	contig	of	7454	bp	in	length
63694	63793:	gap	of	unknown	length		
63794	68500:	contig	of	4707	bp	in	length
68501	68600:	gap	of	unknown	length		
68601	75767:	contig	of	7167	bp	in	length
75768	75867:	gap	of	unknown	length		
75868	82947:	contig	of	7080	bp	in	length

```

* 82948      83047: gap of unknown length
* 83048      89217: contig of 6170 bp in length
* 89317      89317: gap of unknown length
* 89318      96269: contig of 6952 bp in length
* 96270      96369: gap of unknown length
* 96370      102863: contig of 6496 bp in length
* 102866     102965: gap of unknown length
* 102966     113573: contig of 10608 bp in length
* 113574     113673: gap of unknown length
* 113674     121561: contig of 7888 bp in length
* 121562     121661: gap of unknown length
* 121662     131831: contig of 10170 bp in length
* 131832     131931: gap of unknown length
* 131932     143351: contig of 11420 bp in length
* 143352     143351: gap of unknown length
* 143452     154592: contig of 11141 bp in length
* 154593     154692: gap of unknown length
* 154693     165992: contig of 11300 bp in length
* 165993     166092: gap of unknown length
* 166093     189174: contig of 23082 bp in length
* 189175     189274: gap of unknown length
* 189275     230254: contig of 40980 bp in length.
              Location/Qualifiers
                1. .230254
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /clone="RP23-41P9"
                  /clone_lib="RPCI mouse BAC library 23"
61269 a 52644 C 52155 g 60320 t 3966 others
ORIGIN

```

	Query Match	9.4%;	Score 44.4;	DB 2;	Length 230254;
	Best Local Similarity	54.2%;	Pred. No. 4.1;		
	Matches	90;	Conservative	0;	Mismatches 76; Indels 0; Gaps 0;
QY	268	gcgggcttggtcggtttccaaagccgcctagaggctaccgggcgaggggcgcgcgcgaac	327		
Db	40774	GGGGGTTTGGCCGTTGGCCGGGGGGGGGGCCCTCCCCGGGGCGTGGGGGGGCCCTGCGGG	40833		
QY	328	tgcgcgttgcttggtttaccagagacacgtgcgcagtcctccggaagcgcgcgcggggaagac	387		
Db	40834	CGCGCGGGCGGTGGTGCGGGCGCCACGGGGGGCGGCGGTGCCGCCCCCGTTTGG	40893		
QY	388	tgctccgcgcgcgtaccgaggaagcgcgcgcgggtccgcctctgc	433		
Db	40894	GGCTCCCCCGGGCTCCGGCCGCTGGGGGTGGCGGGCCCCCGCCGCG	40939		

RESULT	13
SCD78	
LOCUS	36224 bp DNA linear BCT 26-NOV-1998
DEFINITION	Streptomyces coelicolor cosmid D78.
ACCESSION	AL034355
VERSION	AL034355.1 GI:3928709
KEYWORDS	ABC transporter; ftsK/spoIIIE family protein; hcr2.

KEYWORDS ABC transporter; ftsK/spoIIIE family protein; hsc2; hisdlinol-phosphate aminotransferase; integrase/recombinase; integrated element; oxidoreductase; partial thiamin biosynthesis protein thic; phea; plasmid replication initiator protein; prephenate dehydratase; secreted peptidase; serine/threonine phosphatase; sers; seryl-tRNA synthase; thic'; two component sensor.

SOURCE ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE
<i>Streptomyces coelicolor</i> A3(2).	1	Redenbach, M., Kieser, H.M., Denapafte, D., Eichner, A., Cullum, J., Kinash, H., and Hopwood, D.A.	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb <i>Streptomyces coelicolor</i> A3(2) chromosome	Mol. Microbiol. 21 (1), 77-96 (1996)	97000351
<i>Streptomyces coelicolor</i> A3(2)	1				
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.					
1 (bases 1 to 36224)					

SRIRGQSVSAWRFQWEDDKVTLRPDDEAIAITAVHDVITAGASISSEITRRWKEAGFI
TSRGNPFOYOTVKNVNLNARICGVREIKGEIVRGDDQDQPIVGEWEAIVTPKQMFAYTA
KIRRGHCTGTPRGGLVHKVLLNLRGCVLEDGTVCNNKMGIGIKANDWLKQYHAYM
CKTVDDGCKNTYARGDKTKIIEELVIAKLEDAATKAQDVPDOKAEALERALQSR
RELERNHDEDDIDDEAFRNLPLVLERRIKELRDVOKAEALKAEEAEADIRKS
WGAKLTOKREAMKKVILGAVIALPGGKGNKTFDPLLKPVWKTSE"
4444..4575
/gene="SCD78.05"
/note="thic",
4444..4575
/partial
/gene="SCD78.05"
/note="SCD78.05, thic", probable thiamin biosynthesis
protein thic, partial CDS, len: 43 aa; similar to the
extreme C-terminus of e.g. THIC_ECOLI thiamin biosynthesis
protein thic (631 aa), fasta scores; opt: 97 z-score:
137.7 E(); 2, 46.2% identity in 39 aa overlap and
THIC_MYCTU (547 aa), fasta scores; opt: 143 z-score: 272.4
E(); 6.4e-08, 53.1% identity in 49 aa overlap. Note that
this gene appears to have been interrupted by the
integration of the upstream element"
/codon_start=1
/transl_table=11
/product="putative thiamin biosynthesis protein thic"
/protein_id="CAA22210.1"
/db_xref="GI:3928714"
/translation="SMKISQIRRHQHGSSKGBIEBGMQAKSKEFAAGNRVYLPAD"
4655..6103
/gene="SCD78.06"
4655..6103
/gene="SCD78.06"
/note="SCD78.06, unknown, len: 482 aa; similar to

gene

CDS

gene

CDS

Query Match

Best Local Similarity 9.3%; Score 44; DB 1; Length 36224;
Matches 86; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 262 tcgcccgcgagccttgctggtttccgagccgctagagctaccggcgagggcggggc 321
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18614 TCGCCGACGCCGCGTTCGTGGAGACGCCGCCGCCACCGCCCTGGAACGCCGCC 18673
QY 322 cggagctcgcctgttcgctgttaccagagacagctgcgcagtcgccgaagcgccgg 381
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18674 CCGCGCGCTCTACGCCCGCGTCTGCTGCTGCCCTGTGCCGCGGACGGAGGTCCTCG 18733
QY 382 ggaagctcgc 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18734 GGAGCCTGCTCATGCCACGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 18769

RESULT 14

AC027682/c
LOCUS AC027682 205914 bp DNA linear HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2012K14, WORKING DRAFT
SEQUENCE, 67 unordered pieces.
ACCESSION AC027682
VERSION AC027682.5 GI:13786405
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 205914)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 205914)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:10281392.
-----Genome Center

Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 0
Center clone name: CITB-H1_2012K14

Summary Statistics
Consensus quality: 111602 bases at least Q40
Consensus quality: 122770 bases at least Q30
Estimated insert size: 164311; agarose-fp estimation
Estimated insert size: 199314; sum-of-contigs estimation
Quality coverage: 5.85 in Q20 bases; agarose-fp estimation
Quality coverage: 4.83 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1209 1208: contig of 1208 bp in length
1309 1308: gap of unknown length
2502 2502: contig of 1194 bp in length
2503 2602: gap of unknown length
2603 3803: contig of 1201 bp in length
3804 3903: gap of unknown length
3904 5055: contig of 1152 bp in length
5056 5155: gap of unknown length
5156 6518: contig of 1363 bp in length
6519 7745: gap of unknown length
7746 7845: gap of unknown length
7846 9101: contig of 1256 bp in length
9102 9201: gap of unknown length
9202 10396: contig of 1195 bp in length
10397 10496: gap of unknown length
10497 11697: contig of 1201 bp in length
11698 11798: gap of unknown length
11799 13030: contig of 1233 bp in length
13031 14324: contig of 1194 bp in length
14325 14424: gap of unknown length
14425 15754: gap of unknown length
15754 17109: contig of 1256 bp in length
17110 17209: gap of unknown length
17210 18342: contig of 1133 bp in length
18343 18442: gap of unknown length
18443 19741: contig of 1299 bp in length
19742 19841: gap of unknown length
19842 20944: contig of 1102 bp in length
20944 21043: gap of unknown length
21044 22227: contig of 1184 bp in length
22228 22327: gap of unknown length
22328 23517: contig of 1190 bp in length
23518 23617: gap of unknown length
23618 24857: contig of 1240 bp in length
24858 24957: gap of unknown length
24958 26205: contig of 1248 bp in length
26206 26305: gap of unknown length
26306 27547: contig of 1242 bp in length
27548 27647: gap of unknown length
27648 28909: contig of 1262 bp in length
28910 29010: gap of unknown length
29010 30081: contig of 1072 bp in length
30082 30181: gap of unknown length
30182 31298: contig of 1117 bp in length
31299 31398: gap of unknown length
31399 32604: contig of 1206 bp in length

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 21:53:03 ; Search time 525.6 Seconds
(without alignments)
81.664 Million cell updates/sec

Title: US-09-696-686-45

Perfect score: 25
Sequence: 1 agctcagacatggactccatggccc 25

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	22	AAS05280
2	25	100.0	25	22	AAS05287
3	25	100.0	49	22	AAS05282
4	25	100.0	471	22	AAS05284
5	25	100.0	1839	22	AAS05278
6	25	100.0	1848	22	AAS05286
7	19.2	76.8	1675	21	AAS21311
8	18.6	74.4	435	21	AAZ27030
9	18.6	74.4	585	22	AAS26338

C 10	18.6	74.4	933	22	AAS26339	Human cDNA encodin
C 11	18.6	74.4	1422	21	AAA27028	Human cbl-SL cDNA
C 12	18.6	74.4	1522	22	AAS25883	Human cDNA encodin
C 13	18.6	74.4	1526	22	AAK51619	Human polynucleoti
C 14	18.6	74.4	1543	22	AAK52603	Human polynucleoti
C 15	18.6	74.4	1547	21	AAK27027	Human cbl-SL cDNA.
C 16	18.6	74.4	2964	23	AAS89487	DNA encoding novel
C 17	18.2	72.8	1291	21	AAC49165	Arabidopsis thalia
C 18	18.2	72.8	1295	21	AAC35699	Arabidopsis thalia
C 19	18.2	72.8	30598	23	ABL01908	Drosophila melanog
C 20	17.8	71.2	2384	22	ABD21727	Mouse stem cell gr
C 21	17.6	70.4	54548	21	AAZ45596	DNA sequence of th
C 22	17.6	70.4	157	22	ABA74959	Human foetal liver
C 23	17.6	70.4	157	22	ABA39649	Human foetal liver
C 24	17.6	70.4	157	22	AAK23460	Human brain expres
C 25	17.6	70.4	157	22	AAK49606	Human bone marrow
C 26	17.6	70.4	157	22	AAI26712	Probe #16645 for g
C 27	17.6	70.4	157	22	AAI55487	Probe #24173 used
C 28	17.6	70.4	583	22	ABA62434	Human foetal liver
C 29	17.6	70.4	583	22	ABA29768	Probe #8234 for ge
C 30	17.6	70.4	583	22	AAK10776	Human brain expres
C 31	17.6	70.4	583	22	AAK36652	Human bone marrow
C 32	17.6	70.4	583	22	AAI17503	Probe #7436 for ge
C 33	17.6	70.4	583	22	AAI42414	Probe #11100 used
C 34	17.6	70.4	873	22	AAH46902	cDNA encoding huma
C 35	17.6	70.4	1118	23	AAS85770	DNA encoding novel
C 36	17.6	70.4	1292	14	AAQ47726	Coding sequence of
C 37	17.6	70.4	1327	24	AAS94754	Human DNA sequence
C 38	17.6	70.4	1362	22	AAS05279	Human trinucleotid
C 39	17.6	70.4	1462	20	AAZ24833	Human secreted pro
C 40	17.6	70.4	1478	24	ABJ99729	Mouse ischaemic co
C 41	17.6	70.4	1515	21	AAA87711	Secreted protein e
C 42	17.6	70.4	1515	21	AAA87748	Human secreted pro
C 43	17.6	70.4	1515	22	AAF63993	cDNA encoding secr
C 44	17.6	70.4	1515	22	AAF64030	cDNA encoding huma
C 45	17.6	70.4	1532	21	AAZ64995	Membrane-bound pro

ALIGNMENTS

RESULT 1
AAS05280
ID AAS05280 standard; DNA; 25 BP.
XX AAS05280;
AC AAS05280;
XX 07-SEP-2001 (first entry)
XX PCR primer #1 used to amplify T243 homologous sequences.
XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
XX transgenic animal; knockout mouse; triplet repeat expansion;
XX fragile X syndrome; Huntington's disease; mouse; PCR primer; sa.
XX Mus sp.
XX WO200130798-A1.
XX 03-MAY-2001.
XX 26-OCT-2000; 2000WO-US29382.
XX 26-OCT-1999; 99US-0161488.
XX (DELT-) DELTAGEN INC.
XX Klein R, Matthews W, Moore M, Allen KD;
XX WPI; 2001-300473/31.
XX Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
PT

PT contains heterozygous disruption in a gene encoding TRP -
 XX Claim 13; Fig 13; 106pp; English.
 CC The present sequence for PCR primer #1 is used to amplify sequences
 CC homologous to the T243 gene which encodes for a trinucleotide repeat
 CC protein (TRP). The invention describes methods of producing embryonic
 CC stem (ES) cells comprising a heterozygous disruption in a target DNA
 CC sequence (preferably gene T243) encoding a TRP and of producing a
 CC knockout mouse comprising a homozygous disruption in a gene encoding TRP,
 CC where the disruption inhibits the production of the wild type TRP. The
 CC invention also relates to identifying agents capable of affecting a
 CC phenotype of a knockout mouse. Also described are methods of determining
 CC whether expansion of the trinucleotide repeat in a gene encoding TRP
 CC produces a phenotypic change. The transgenic animals and the cells are
 CC useful for identifying compounds capable of ameliorating disease
 CC symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.
 XX
 SQ Sequence 25 BP; 6 A; 9 C; 6 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.066;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 agctcagacatgactccatgccc 25
 Db 1 agctcagacatgactccatgccc 25

RESULT 2
 AAS05287/C
 ID AAS05287 standard; DNA; 25 BP.
 XX
 AC AAS05287;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Mouse ES cell T243 gene PCR primer #426.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW mouse; PCR primer; ss.
 XX
 OS Mus sp.
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELT-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPI; 2001-300473/31.
 XX
 DR Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP -
 XX
 PS Example 12; Fig 13; 106pp; English.

XX The present sequence for PCR primer #426 is used to amplify the T243
 CC gene from an R1 mouse embryonic stem cell genomic library. The T243 gene
 CC encodes for a trinucleotide repeat protein (TRP). The invention describes
 CC methods of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably T243) encoding a TRP and
 CC of producing a knockout mouse comprising a homozygous disruption in a
 CC gene encoding TRP, where the disruption inhibits the production of the
 CC wild type TRP. The invention also relates to identifying agents capable
 CC of affecting a phenotype of a knockout mouse. Also described are methods
 CC of determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and
 CC the cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.
 XX
 SQ Sequence 25 BP; 4 A; 6 C; 9 G; 6 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;
 Best Local Similarity 100.0%; Pred. No. 0.066;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 agctcagacatgactccatgccc 25
 Db 25 agctcagacatgactccatgccc 1

RESULT 3
 AAS05282
 ID AAS05282 standard; DNA; 49 BP.
 XX
 AC AAS05282;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE PCR primer #488 used to amplify T243 PCR product from pool H10.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; mouse; PCR primer; ss.
 XX
 OS Mus sp.
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELT-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPI; 2001-300473/31.
 XX
 DR Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP -
 XX
 PS Example 12; Fig 13; 106pp; English.
 XX
 CC The present sequence for PCR primer #488 is used to further amplify
 CC a T243 gene PCR product from pool H10 (obtained by PCR of the T243 gene

CC from an R1 mouse embryonic stem cell genomic library). Gene T243 encodes
 CC for a trinucleotide repeat protein (TRP). The invention describes methods
 CC of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably gene T243) encoding a
 CC TRP and of producing a knockout mouse comprising a homozygous disruption
 CC in a gene encoding TRP. The invention also relates to identifying agents
 CC capable of affecting a phenotype of a knockout mouse. The transgenic
 CC animals and the cells are useful for identifying compounds capable of
 CC ameliorating disease symptoms, and as test substrates for the
 CC identification of drugs, pharmaceuticals, therapies and interventions
 CC which may be effective in treating trinucleotide repeat disorders e.g.
 CC fragile X syndrome and Huntington's disease. The animal models for
 CC trinucleotide repeat disorders are ideal model systems to study the
 CC progression of disease in vivo, the molecular basis of these diseases
 CC and show the features observed in human disease. Using the mice, it is
 CC possible to model both the pathogenic mechanism and the trinucleotide
 CC repeat instability in the mouse.

XX
 SQ Sequence 49 BP; 8 A; 16 C; 13 G; 12 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 49;
 Best Local Similarity 100.0%; Pred. No. 0.071;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||||
 Db 25 agctcagacatggactccatggccc 49

RESULT 4
 AAS05284/c
 ID AAS05284 standard; DNA: 471 BP.

XX AC AAS05284;

XX DT 07-SEP-2001 (first entry)

XX DE T243 gene homologous sequence #1 generated by PCR.

XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 XX transgenic animal; knockout mouse; triplet repeat expansion;
 XX fragile X syndrome; Huntington's disease; mouse; ds.

XX OS Mus sp.

XX PN WO200130798-A1.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-US29382.

XX PR 26-OCT-1999; 99US-0161488.

XX PA (DELT-) DELTAGEN INC.

XX PI Klein R, Matthews W, Moore M, Allen KD;

XX DR WPI; 2001-300473/31.

XX PT Novel transgenic animals useful as animal model for characterization of
 XX function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -

XX PS Claim 11; Fig 14; 106pp; English.

XX CC The present sequence for T243 homologous sequence #1 is generated by PCR
 XX and is homologous to the T243 gene which encodes for a trinucleotide
 XX repeat protein (TRP). The invention describes methods of producing
 XX embryonic stem (ES) cells comprising a heterozygous disruption in a
 XX target DNA sequence (preferably gene T243) encoding a TRP and of
 XX producing a knockout mouse comprising a homozygous disruption in a gene

CC encoding TRP, where the disruption inhibits the production of the wild
 CC type TRP. The invention also relates to identifying agents capable of
 CC affecting a phenotype of a knockout mouse. Also described are methods of
 CC determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and the
 CC cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.

XX SQ Sequence 471 BP; 105 A; 129 C; 148 G; 88 T; 1 other;

Query Match 100.0%; Score 25; DB 22; Length 471;
 Best Local Similarity 100.0%; Pred. No. 0.092;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
 |||||
 Db 471 AGCTCAGACATGGACTCCATGGCCC 447

RESULT 5

AAS05278/c
 ID AAS05278 standard; cDNA: 1839 BP.

XX AC AAS05278;

XX DT 07-SEP-2001 (first entry)

XX DE Murine trinucleotide repeat protein (TRP) cDNA sequence.

XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
 XX transgenic animal; knockout mouse; triplet repeat expansion;
 XX fragile X syndrome; Huntington's disease; mouse; murine; ss.

XX OS Mus sp.

XX FH Key Location/Qualifiers

XX FT CDS 55..885

XX FT /*tag= a

XX FT /product= "TRP"

XX PN WO200130798-A1.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-US29382.

XX PR 26-OCT-1999; 99US-0161488.

XX PA (DELT-) DELTAGEN INC.

XX PI Klein R, Matthews W, Moore M, Allen KD;

XX XX WPI; 2001-300473/31.

XX DR P-PSDB; AAU02498.

XX PT Novel transgenic animals useful as animal model for characterization of
 XX function of a gene encoding trinucleotide repeat proteins (TRPs),
 XX contains heterozygous disruption in a gene encoding TRP -

XX PS Claim 8; Fig 11; 106pp; English.

XX CC The present sequence encodes for murine trinucleotide repeat protein
 XX (TRP). The invention describes methods of producing embryonic stem (ES)
 XX cells comprising a heterozygous disruption in a target DNA sequence

Fri May 24 10:49:36 2002

us-09-696-686-45.rng

CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
CC comprising a homozygous disruption in a gene encoding TRP, where the
CC disruption inhibits the production of the wild type TRP. The invention
CC also relates to identifying agents capable of affecting a phenotype of
CC a knockout mouse. Also described are methods of determining whether
CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
CC phenotypic change. The transgenic animals and the cells are useful for
CC identifying compounds capable of ameliorating disease symptoms, and as
CC test substrates for the identification of drugs, pharmaceuticals,
CC therapies and interventions which may be effective in treating
CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
CC disease. The animal models for trinucleotide repeat disorders are ideal
CC model systems to study the progression of disease in vivo, the molecular
CC basis of these diseases and show the features observed in human disease.
CC using the mice, it is possible to model both the pathogenic mechanism and
CC the trinucleotide repeat instability in the mouse.
XX
SQ Sequence 1839 BP; 424 A; 522 C; 535 G; 358 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 1839;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agctcagacatggactccatggccc 25
|||||
Db 74 AGCTCAGACATGGACTCCATGGCCC 50

RESULT 6
AAS05286/c
ID AAS05286 standard; DNA; 1848 BP.
XX
AC AAS05286;
XX
DT 07-SEP-2001 (first entry)
XX
DE DNA sequence of an expanded T243 gene.
XX
KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES;
KW transgenic animal; knockout mouse; triplet repeat expansion;
KW fragile X syndrome; Huntington's disease; mouse; mutant; ds.
XX
OS Mus sp.
OS Synthetic.

XX Key Location/Qualifiers
FH 55..894
FT /*tag= a
FT /product= "Expanded TRP"
XX
XX WO200130798-A1.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-US29382.
XX
XX 26-OCT-1999; 99US-0161488.
XX
XX (DELT-) DELTAGEN INC.
XX
XX Klein R, Matthews W, Moore M, Allen KD;
PI WPI; 2001-300473/31.
DR P-PSDB; AAU02500.
XX
XX Novel transgenic animals useful as animal model for characterization of
PT function of a gene encoding trinucleotide repeat proteins (TRPs),
PT contains heterozygous disruption in a gene encoding TRP -
XX
XX Disclosure; Fig 15; 106pp; English.
PS
XX The present sequence represents the DNA sequence of an expanded T243
CC

CC gene. The T243 gene encodes, for a trinucleotide repeat protein (TRP).
CC The invention describes methods of producing embryonic stem (ES)
CC cells comprising a heterozygous disruption in a target DNA sequence
CC (preferably gene T243) encoding a TRP and of producing a knockout mouse
CC comprising a homozygous disruption in a gene encoding TRP, where the
CC disruption inhibits the production of the wild type TRP. The invention
CC also relates to identifying agents capable of affecting a phenotype of
CC a knockout mouse. Also described are methods of determining whether
CC expansion of the trinucleotide repeat in a gene encoding TRP produces a
CC phenotypic change. The transgenic animals and the cells are useful for
CC identifying compounds capable of ameliorating disease symptoms, and as
CC test substrates for the identification of drugs, pharmaceuticals,
CC therapies and interventions which may be effective in treating
CC trinucleotide repeat disorders e.g. fragile X syndrome and Huntington's
CC disease. The animal models for trinucleotide repeat disorders are ideal
CC model systems to study the progression of disease in vivo, the molecular
CC basis of these diseases and show the features observed in human disease.
CC using the mice, it is possible to model both the pathogenic mechanism and
CC the trinucleotide repeat instability in the mouse.
XX
SQ Sequence 1848 BP; 415 A; 519 C; 543 G; 371 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 1848;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agctcagacatggactccatggccc 25
|||||
Db 74 AGCTCAGACATGGACTCCATGGCCC 50

RESULT 7
AAS21311/c
ID AAS21311 standard; cDNA; 1675 BP.
XX
AC AAS21311;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human cDNA sequence encoding for PRO4409 polypeptide.
XX
KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200140466-A2.
XX
XX 07-JUN-2001.
XX
XX 01-DEC-2000; 2000WO-US32678.
XX
XX 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 09-DEC-1999; 99US-0170262.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31243.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00276.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.

PA (GETH) GENENTECH INC.

PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

DR WPI: 2001-408281/43.
 DR P-PSDB: AAU12239.

XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical

PS Claim 3; Fig 135; 813pp; English.

XX AAS21244-AAS21518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.

XX Sequence 1675 BP; 380 A; 469 C; 520 G; 306 T; 0 other;

Query Match 76.8%; Score 19.2; DB 22; Length 1675;
 Best Local Similarity 87.5%; Pred. No. 39;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gctcagacatggactccatgcc 25
 Db 355 GCTCAGCATGGACTCCATGCC 332
 ||||| ||| ||||| |||||

RESULT 8
 AAA27030/c
 ID AAA27030 standard; cDNA; 435 BP.

XX AAA27030;

XX 22-AUG-2000 (first entry)

XX Human EST with GenBank Accession No. AAL12513.

XX Human; cbl-SL; proto-oncogene; tyrosine kinase activity modulation;
 KW cancer; carcinoma; cell proliferation; cell differentiation;
 KW cytostatic; gene therapy; EST; expressed sequence tag; ss.

OS Homo sapiens.
 XX WO200027865-A1.
 PN 18-MAY-2000.

PD 05-NOV-1999; 99WO-US26057.

PF 06-NOV-1998; 98US-0107470.

PR (BGHM) BRIGHAM & WOMENS HOSPITAL INC.

XX Borriello F, Band H;

XX WPI: 2000-376482/32.

XX Novel nucleic acids encoding cbl-SL proteins useful for treating
 PT cancers expressing a mutant cbl-SL polypeptide

PS Claim 21; Page 71; 77pp; English.

XX The present sequence is a human EST with Genbank Accession Number
 CC AAL12513. It shows partial homology to the coding sequence for the human
 CC protein cbl-SL (Y94330), which is believed to play a role in the
 CC regulation of a cell's growth, differentiation and proliferation.
 CC Aberrant expression levels and mutant forms of tyrosine kinases which are
 CC the likely targets of cbl-SL have been found in patients with certain
 CC types of cancer. The cbl-SL nucleotide sequence was isolated by screening
 CC a human pancreatic adenocarcinoma cDNA library using the insert of
 CC I.M.A.G.E clone ID 526956 from ATCC. Individual cDNA clones were
 CC subjected to 3-4 rounds of amplification and purification and several
 CC inserts were sequenced. The sequence of the resulting cbl-SL full-length
 CC clone was confirmed on more independent clones. The nucleic acid sequence
 CC may be used to screen for the presence of a carcinoma in a subject by
 CC comparing the cbl-SL DNA or mRNA sequence in the test sample to the
 CC cbl-SL DNA or mRNA of a control sample. Levels of cbl-1 expression may
 CC be measured using PCR and Northern blotting. The nucleic acid sequence
 CC may also be used in gene therapy and in the construction of a cbl-SL gene
 CC knockout in cells and animals used to study cbl-SL activity. cbl-SL
 CC polypeptides can be used to screen peptide libraries, including phage
 CC display libraries, to identify and select peptide binding partners of the
 CC cbl-SL protein. These can be used for screening assays, for purification
 CC protocols and for interfering directly with the functioning of cbl-SL.
 CC Any binding agents that also bind to the present sequence are excluded
 CC from the invention. The cbl-SL protein may also be administered as a
 CC therapeutic agent to cancer patients expressing a mutant cbl-SL.

XX Sequence 435 BP; 94 A; 136 C; 121 G; 82 T; 2 other;

Query Match 74.4%; Score 18.6; DB 21; Length 435;
 Best Local Similarity 84.0%; Pred. No. 63;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatgcc 25
 Db 287 AGCTCAAAATGGAGTCCATGCC 263
 ||||| ||| ||||| |||||

RESULT 9
 AAS26338
 ID AAS26338 standard; cDNA; 585 BP.

XX AAS26338;

XX 07-NOV-2001 (first entry)

XX Human cDNA encoding a novel secreted protein, Seq ID 517.

XX Human; immunosuppressive; antiarthritic; ss; antirheumatic;

KW cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;

KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

KW vulnery; secreted protein; rheumatoid arthritis;

KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin ageing; food additive; preservative; antiproliferative.

OS Homo sapiens.

XX WO200155322-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01341.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.


```

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488783/53.
DR P-PSDB; AAU16351.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 1; SEQ ID No 517; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunoassorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.
XX
XX Query Match 74.48; Score 18.6; DB 22; Length 585;
XX Best Local Similarity 84.08; Pred. No. 65;
XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX Qy 1 agctcagacatggactccatggccc 25
XX ||||| |||| |||||
XX Db 484 agctcaaatgtggatccatggccc 508
XX
XX RESULT 10
XX AAS26339/C
XX ID AAS26339 standard; cDNA; 933 BP.
XX AC AAS26339;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human cDNA encoding a novel secreted protein, Seq ID 518.
XX
XX Human; immunosuppressive; antiarthritic; ss; antirheumatic;
XX cytosolic; cardiol; vasotropic; cerebroprotective; neurotropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnerary; secreted protein; rheumatoid arthritis;
XX hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
XX cerebrovascular disorder; cerebral ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; infection; ocular disorder;
XX corneal infection; wound healing; epithelial cell proliferation;
XX skin aging; food additive; preservative; antiproliferative.
XX
XX Homo sapiens.
XX

```

```

XX WO200155322-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01341.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220984.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.

```

27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251388.
05-DEC-2000; 2000US-0256179.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251857.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
PA XX
XX XX

PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-488783/53;
DR P-PSDB; AAU16352.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
XX Claim 1; SEQ ID No 518; 980pp; English.
PS
XX The invention relates to isolated nucleic acid molecules and their
XX encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC cardiovascular disorders e.g. cardiac arrest, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

Query Match 74.4%; Score 18.6; DB 22; Length 933;
Best Local Similarity 84.0%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY: 1 agctcagacatggactccatggccc 25
||||| ||||| ||||| ||||| |||||
DB 447 AGCTCAATGTGGAGTCCCATGCCCC 423

RESULT 11
AAA27028/c
ID AAA27028 standard; cDNA; 1422 BP.
XX
AC AAA27028;
XX
DT 22-AUG-2000 (first entry)
XX
DE Human cbl-SL cDNA largest open reading frame.
KW Human; proto-oncogene; tyrosine kinase activity modulation;
KW cancer; carcinoma; cell proliferation; cell differentiation;
KW cytostatic; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1422
FT /*tag= a
FT /product= "cbl-SL"
XX
XX WO200027865-A1.
XX
XX 18-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26057.
PF
XX 06-NOV-1998; 98US-0107470.
PR

XX PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX PI Borriello F, Band H;
XX PF WPI: 2000-376482/32.
DR DR P-PSDB: AAY94330.
XX
PT Novel nucleic acids encoding cbl-SL proteins useful for treating
PT cancers expressing a mutant cbl-SL polypeptide
XX
XX Claim 3; Page 70-71; 77pp; English.
XX
CC The present sequence is the largest open reading frame of the human
CC cbl-SL cDNA (AAA27027) which encodes the human protein cbl-SL. cbl-SL
CC is believed to play a role in the regulation of a cell's growth,
CC differentiation and proliferation. Aberrant expression levels and mutant
CC forms of tyrosine kinases which are the likely targets of cbl-SL have
CC been found in patients with certain types of cancer. The sequence was
CC isolated by screening a human pancreatic adenocarcinoma cDNA library
CC using the insert of I.M.A.G. E clone ID 526956 from ATCC. Individual cDNA
CC clones were subjected to 3-4 rounds of amplification and purification
CC and several inserts were sequenced. The sequence of the resulting cbl-SL
CC full-length clone was confirmed on more independent clones. The
CC nucleic acid sequence may be used to screen for the presence of a
CC carcinoma in a subject by comparing the cbl-SL DNA or mRNA sequence in
CC the test sample to the cbl-SL DNA or mRNA of a control sample. Levels of
CC cbl-1 expression may be measured using PCR and Northern blotting. The
CC nucleic acid sequence may also be used in gene therapy and in the
CC construction of a cbl-SL gene knockout in cells and animals used to
CC study cbl-SL activity. cbl-SL polypeptides can be used to screen peptide
CC libraries, including phage display libraries, to identify and select
CC peptide binding partners of the cbl-SL protein. These can be used
CC for screening assays, for purification protocols and for interfering
CC directly with the functioning of cbl-SL. The protein may also
CC be administered as a therapeutic agent to cancer patients expressing a
CC mutant cbl-SL.
XX
SQ Sequence 1422 BP; 269 A; 484 C; 432 G; 237 T; 0 other;

Query Match 74.4%; Score 18.6; DB 21; Length 1422;
Best Local Similarity 84.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
||||| ||| |||||
DB 1049 AGCTCAAAATGTGGAGTCATGCGCC 1025

RESULT 12
AAS25883/C
ID AAS25883 standard; cDNA; 1522 BP.
XX AC
XX AC AAS25883;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human cDNA encoding a novel secreted protein, Seq ID 62.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cystostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN W0200155322-A2.

XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01341.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225286.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.
XX 26-SEP-2000; 2000US-0235484.
XX 27-SEP-2000; 2000US-0235834.
XX 27-SEP-2000; 2000US-0235836.
XX 29-SEP-2000; 2000US-0236327.


```
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR P-PSDB: AAM78486.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 1; Page 889-891; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 1526 BP; 303 A; 505 C; 465 G; 253 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 1526;
Best Local Similarity 84.0%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 agctcagacatggactccatggccc 25
Db 1049 AGCTCAAAATGTGGAGTCCATGCC 1025

RESULT 14
AAK52603/C
ID AAK52603 standard; cDNA: 1543 BP.
XX AC AAK52603;
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 2132.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
```

```
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR P-PSDB: AAM79470.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 1; Page 4504; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 1543 BP; 306 A; 515 C; 470 G; 252 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 1543;
Best Local Similarity 84.0%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 agctcagacatggactccatggccc 25
Db 1066 AGCTCAAAATGTGGAGTCCATGCC 1042

RESULT 15
AAA27027/C
ID AAA27027 standard; cDNA: 1547 BP.
XX AC AAA27027;
XX DT 22-AUG-2000 (first entry)
XX DE Human cbl-SL cDNA.
XX KW Human; proto-oncogene; tyrosine Kinase activity modulation;
XX KW cancer; carcinoma; cell proliferation; cell differentiation;
XX KW cytostatic; gene therapy; ss.
XX OS Homo sapiens.
XX PN WO200027865-A1.
XX PD 18-MAY-2000.
XX PF 05-NOV-1999; 99WO-US26057.
XX PR 06-NOV-1998; 98US-0107470.
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
```


CLASSIFICATION: 435
PRIOR APPLICATION DATA: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-913-805A-9

Query Match 65.6%; Score 16.4; DB 3; Length 2100;
Best Local Similarity 94.4%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 gctcagacatggactcca 19
Db 1294 GCTCTGACATGGACTCCA 1277

RESULT 15
US-09-442-629-9/c
Sequence 9, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
SHIBA, Tetsuo
YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442.629
FILING DATE: 18-NO. 6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913.805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996

APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-442-629-9

Query Match 65.6%; Score 16.4; DB 4; Length 2100;
Best Local Similarity 94.4%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 gctcagacatggactcca 19
Db 1294 GCTCTGACATGGACTCCA 1277

Search completed: May 24, 2002, 00:18:05
Job time: 11551 sec

```

; NAME/KEY: CDS
; LOCATION: (757)..(3771)
US-08-935-450-1

Query Match 67.2%; Score 16.8; DB 2; Length 4534;
Best Local Similarity 90.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 cagacatggactccatggcc 24
||||| ||||| ||||| |||||
Db 2686 CAGATGGTCTCCATGCC 2667

RESULT 13
US-09-239-909-3
; Sequence 3, Application US/09239909
; Patent No. 6284952
; GENERAL INFORMATION:
; APPLICANT: Kumho Petrochemical Co. Ltd.
; TITLE OF INVENTION: Transgenic Plants with Divergent ScaM4 or ScaM5 Gene to
; FILE REFERENCE: P99-2-6
; CURRENT APPLICATION NUMBER: US/09/239,909
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: EP 99300136.1
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: KOPATIN 1.0
; SEQ ID NO 3
; LENGTH: 916
; TYPE: DNA
; ORGANISM: G. max calmodulin5 (ScaM5)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (69)..(518)
US-09-239-909-3

Query Match 56.4%; Score 16.6; DB 4; Length 916;
Best Local Similarity 82.6%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggc 23
||||| ||||| || |||||
Db 51 agctcaaacattaacaccatggc 73

RESULT 14
US-08-913-805A-9/C
; Sequence 9, Application US/08913805A
; Patent No. 6054304
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; SYMPT: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998

```


Qy 1 agctcagacatggactccat 20

RESULT 11
US-08-888-077A-41
; Sequence 41, Application US/08888077A
; Patent No. 6020143
; GENERAL INFORMATION:

```

? APPLICATION NUMBER: US/09/385, 028
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? PRIORITY NUMBER: US 08/790,462
? FILING DATE: 29-JAN-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: D. Douglas Price
? REGISTRATION NUMBER: 24,514
? REFERENCE/DOCKET NUMBER: 1418/P57452US2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 638-6666
? TELEFAX: (202) 39305350
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 11604 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-09-385-028-13

Query Match 68.0%; Score 17; DB 4; Length 11604;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 agctcagacatggactccatgcccc 25
||||| ||||| ||||| ||||| |||||
DB 1591 AGCTCCGACCTGGAGCATTGCC 1615

RESULT 8
US-09-385-028-1
? Sequence 1, Application US/09385028
? Patent No. 6232106
? GENERAL INFORMATION:
? APPLICANT: Susan E. Jensen
? APPLICANT: Kwamena A Aidoo
? APPLICANT: Ashish S. Paradkar
? TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
? Patent No. 6232106
? TITLE OF INVENTION: Acid Biosynthesis
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN,PPLC
? STREET: The Jenifer Building, 400 Seventh Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20004
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/385,028
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/790,462
? FILING DATE: 29-JAN-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: D. Douglas Price
? REGISTRATION NUMBER: 24,514
? REFERENCE/DOCKET NUMBER: 1418/P57452US2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 638-6666
? TELEFAX: (202) 39305350
? TELEX: RCA 248593 IDEA UR
? INFORMATION FOR SEQ ID NO: 1:

```

Query Match 70.4%; Score 17.6; DB 4; Length 7720;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gctcagacatggactccatggccc 25
Db 3631 gctcagacaggattgcatggccc 3654

RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 68.8%; Score 17.2; DB 4; Length 4403765;
Best Local Similarity 86.4%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gctcagacatggactccatggccc 23
Db 1903992 GCACAGACATGATCCCATGGC 1903971

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 68.8%; Score 17.2; DB 4; Length 4411529;
Best Local Similarity 86.4%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gctcagacatggactccatggccc 23
Db 1913221 GCACAGACATGATCCCATGGC 1913200

RESULT 5
US-09-385-028-15
; Sequence 15, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSER: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jenifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-385-028-15

Query Match 68.0%; Score 17; DB 4; Length 1722;
Best Local Similarity 80.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25
Db 1408 AGCTCCGACCTGGAGACCATCGCCC 1432

RESULT 6
US-09-014-969-10
; Sequence 10, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2002, 21:05:34 ; Search time 121.39 Seconds
(without alignments)
50.588 Million cell updates/sec

Title: US-09-696-686-45
Perfect score: 25
Sequence: 1 agctcagacatggactccatggccc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	17.6	70.4	3228	4	US-09-318-448-20
2	17.6	70.4	7720	4	US-09-318-448-5
3	17.2	68.8	4403765	4	US-09-103-840A-2
4	17.2	68.8	4411529	4	US-09-103-840A-1
5	17	68.0	1722	4	US-09-385-028-15
6	17	68.0	2925	2	US-09-014-969-10
7	17	68.0	11604	4	US-09-385-028-13
8	17	68.0	15079	4	US-09-385-028-1
9	16.8	67.2	539	1	US-08-322-742-18
10	16.8	67.2	2036	3	US-08-923-454A-17
11	16.8	67.2	2205	3	US-08-888-077A-41
12	16.8	67.2	4534	2	US-08-935-450-1
13	16.6	66.4	916	4	US-09-239-909-3
14	16.4	65.6	2100	3	US-08-913-805A-9
15	16.4	65.6	2100	4	US-09-442-629-9
16	16.2	64.8	2268	2	US-08-890-853-1
17	16.2	64.8	2268	2	US-09-099-125A-1
18	16.2	64.8	2268	2	US-09-099-124A-1
19	16.2	64.8	2268	2	US-09-197-008-1
20	16.2	64.8	2268	4	US-09-032-476-1
21	16.2	64.8	2268	4	US-08-890-854-1
22	16.2	64.8	2268	4	US-09-023-324-1
23	16.2	64.8	2271	4	US-08-910-820-8
24	16.2	64.8	2931	4	US-09-168-629-14
25	16.2	64.8	3966	3	US-09-215-131-1
26	16.2	64.8	3966	3	US-09-222-734-1
27	16.2	64.8	7886	2	US-08-751-189-2

C 28	16.2	64.8	7886	2	US-09-060-836-2	Sequence 2, Appli
C 29	16.2	64.8	7886	4	US-09-184-445-2	Sequence 2, Appli
C 30	16	64.0	327	4	US-08-836-075A-5	Sequence 5, Appli
C 31	16	64.0	808	3	US-08-851-136C-15	Sequence 15, Appli
C 32	16	64.0	2415	3	US-09-019-689-1	Sequence 1, Appli
C 33	16	64.0	3407	1	US-08-253-155A-7	Sequence 7, Appli
C 34	16	64.0	4374	2	US-08-449-644-3	Sequence 3, Appli
C 35	16	64.0	4374	2	US-08-087-244A-3	Sequence 3, Appli
C 36	16	64.0	11531	1	US-08-068-945A-1	Sequence 1, Appli
C 37	16	64.0	11531	1	US-08-442-806-1	Sequence 1, Appli
C 38	15.8	63.2	1329	2	US-08-758-621-1	Sequence 1, Appli
C 39	15.8	63.2	1329	4	US-09-107-858-1	Sequence 46, Appli
C 40	15.8	63.2	2464	4	US-09-268-992-46	Sequence 48, Appli
C 41	15.8	63.2	2488	4	US-09-268-992-48	Sequence 47, Appli
C 42	15.8	63.2	3016	4	US-09-268-992-47	Sequence 246, Appli
C 43	15.8	63.2	3323	4	US-08-936-165A-246	Sequence 20, Appli
C 44	15.6	62.4	815	4	US-09-541-941B-20	Sequence 7, Appli
C 45	15.6	62.4	877	3	US-08-545-809A-7	

ALIGNMENTS

RESULT 1
US-09-318-448-20
; Sequence 20, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318.448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 3228
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-20

Query Match 70.4%; Score 17.6; DB 4; Length 3228;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gctcagacatggactccatggccc 25
||||| || || ||||| ||
Db 2785 gctcagcaggcctccatggccc 2808

RESULT 2
US-09-318-448-5
; Sequence 5, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318.448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-5

FEATURES High quality sequence stop: 262.
source Location/Qualifiers
1..651
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_lib="IMAGE:3484616"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 163 a 212 c 172 g 104 t
ORIGIN
Query Match 100.0%; Score 25; DB 10; Length 651;
Best Local Similarity 100.0%; Pred. NO. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agctcagacatggactccatggccc 25
|||||
Db 98 AGCTCAGACATGGACTCCATGGCCC 74

Search completed: May 23, 2002, 23:05:27
Job time: 7483 sec

/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCAACTCGAGTCTTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGTCTCGAGTAAATTAATTCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

138 a 162 c 216 g 127 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 643;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agctcagacatggactccatggccc 25
|||||

Db 101 AGCTCAGACATGGACTCCATGGCCC 77

RESULT 14
BB652011/c
LOCUS
DEFINITION BB652011 RIKEN full-length enriched, ES cells Mus musculus cDNA clone C330027E02 5', mRNA sequence.
ACCESSION BB652011
VERSION BB652011.1 GI:15402706
KEYWORDS EST.
ORGANISM house musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Onno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,E., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system---384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
source
1. 649
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C330027E02"
/clone_lib="RIKEN full-length enriched, ES cells"
/cell_type="ES cells"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCAACTCGAGTCTTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGTCTCGAGTAAATTAATTCCTCCCTCCCTCC 3']"

BASE COUNT 135 a 170 c 197 g 147 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 649;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agctcagacatggactccatggccc 25
|||||

Db 94 AGCTCAGACATGGACTCCATGGCCC 70

RESULT 15
BE289316/c
LOCUS
DEFINITION BE289316 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484616 5', mRNA sequence.
ACCESSION BE289316
VERSION BE289316.1 GI:9169606
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM8517 row: n column: 09

,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

TITLE JOURNAL COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse

FEATURES source

Location/Qualifiers
1. .641
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A530052E13"
/clone_lib="RIKEN full-length enriched, adult male aorta
and vein"
/sex="male"
/tissue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGAGAGATTCGAGTTAAATTAATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT ORIGIN

142 a 161 c 206 g 132 t

Query Match 100.0%; Score 25; DB 9; Length 641;
Best Local Similarity 100.0%; Pred. No. 4;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agctcagacatggactccatggccc 25
|||||
Db 67 AGCTCAGACATGGACTCCATGGCCC 43

RESULT 13

BB661603/C

LOCUS BB661603 643 bp mRNA linear EST 26-OCT-2001
DEFINITION BB661603 RIKEN full-length enriched, 0 day neonate kidney Mus
musculus cDNA clone D630047M13 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB661603 BB661603 643 bp mRNA linear EST 26-OCT-2001
BB661603 RIKEN full-length enriched, 0 day neonate kidney Mus
musculus cDNA clone D630047M13 5', mRNA sequence.
BB661603
EST BB661603.1 GI:16495382
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
, Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.

Location/Qualifiers
1. .643
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630047M13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"

FEATURES

source

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

nonredundant cDNA library: Genome Res. 12, 1277-1282 (2002).
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```
prepare mouse tissues:
  Location/Qualifiers
    1. .608
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="E030013D02"
      /clone_lib="RIKEN full-length enriched, 0 day neonate
        lung"
      /tissue_type="lung"
      /dev_stage="0 day neonate"
      /db_xref="europa"
```

Project of Genome Expiration research group at the Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCCAACTCGAGTGTGTGTGTGTGTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence [5', GAGAGAGAGATTCGAGTGAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. "

```

tech 100.0%; Score 25; DB 9; Length 608;
  Similarity 100.0%; Pred. No. 4;
  25; Conservative 0; Mismatches 0; Indels 0; Gaps 0
agotcagacatggactccatggccc 25
|||||
GGCTCAGACATGGACTCCATGGCCC 14

```

BS661026 623 bp mRNA linear EST 26-OCT-2001
 BS661026 RIKEN full-length enriched 0 day neonate kidney Mus
 musculus cDNA clone D630017H14 5', mRNA sequence.
 BS661026
 BS661026.1 GI:16494805
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

TITLE Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
JOURNAL RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
COMMENT Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES

source
1. .485
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G370047K04"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
/cell_type="B16 F10Y cells"
BASE COUNT 101 a 132 c 154 g 98 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 485;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agctcagacatggactccatggccc 25
|||||
Db 92 AGCTCAGACATGGACTCCATGGCCC 68

RESULT 6
BB853356/c
LOCUS BB853356 RIKEN full-length enriched, B16 F10Y cells Mus musculus
DEFINITION cDNA clone G370012D12 5', mRNA sequence.
ACCESSION BB853356
VERSION BB853356.1 GI:7094810
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Salto,R., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa

TITLE
JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES

source
1. .502
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G370012D12"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
/cell_type="B16 F10Y cells"
BASE COUNT 103 a 136 c 161 g 102 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agctcagacatggactccatggccc 25
|||||
Db 89 AGCTCAGACATGGACTCCATGGCCC 65

RESULT 7
BF780486/c
LOCUS BF780486 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:422133
DEFINITION 5', mRNA sequence.
ACCESSION BF780486
VERSION BF780486.1 GI:12085519
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 599)
TITLE NIH-MGC <http://mgs.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-f@mail.nih.gov

JOURNAL COMMENT
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES

Location/Qualifiers
1..472
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G430034D12"
/clone_lib="RIKEN full-length enriched, Leydig cells
CRL-2065 MLTC-1 cDNA"
/cell_type="Leydig cells"
/cell_line="CRL-2065 MLTC-1"
92 a 127 c 156 g 97 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agctcagacatggactccatggccc 25
|||||
Db 96 AGCTCAGACATGGACTCCATGGCCC 72

RESULT 4
BB857917/c
LOCUS BB857917 RIKEN full-length enriched, B16 F10Y cells EST 26-NOV-2001
DEFINITION CDNA clone G370047K11 5', mRNA sequence.
BB857917
ACCESSION BB857917 GI:17099371
VERSION BB857917.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 483)
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

FEATURES

Location/Qualifiers
1..483
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G370047K11"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"
/cell_type="B16 F10Y cells"
95 a 132 c 159 g 97 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 483;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agctcagacatggactccatggccc 25
|||||
Db 112 AGCTCAGACATGGACTCCATGGCCC 88

RESULT 5
BB857913/c
LOCUS BB857913 RIKEN full-length enriched, B16 F10Y cells EST 26-NOV-2001
DEFINITION CDNA clone G370047K04 5', mRNA sequence.
BB857913
ACCESSION BB857913 GI:17099367
VERSION BB857913.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 485)
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

Source

Location/Qualifiers

1..260

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="E330035012"

/clone_lib="RIKEN full-length enriched, 2 days pregnant adult female ovary"

/sex="female"

/tissue_type="ovary"

/dev_stage="2 days pregnant adult"

/lab_host="DH10B"

/note="Site 1: Sali; Site 2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCGCAACTCGAGTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAATTAATTCCTCCCGCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT

ORIGIN

33 a 81 c 88 g 58 t

Query Match

Best Local Similarity 100.0%; Score 25; DB 9; Length 260;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25

|||||

Db 79 ACCTCAGACATGGACTCCATGGCCC 55

RESULT

AA002532/c

LOCUS

DEFINITION

AA002532

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 406)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, N., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE JOURNAL COMMENT

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnlni.gov) for further information.

MGI:262115

Seq primer: mob.REGA+ET

High quality sequence stop: 328.

FEATURES Source

Location/Qualifiers

1..406

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:427563"

/clone_lib="Soares mouse embryo NBME13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCGAATTTTTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru KO, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

81 a 107 c 128 g 90 t

Query Match

Best Local Similarity 100.0%; Score 25; DB 9; Length 406;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agctcagacatggactccatggccc 25

|||||

Db 44 ACCTCAGACATGGACTCCATGGCCC 20

RESULT

BB863087/c

LOCUS

DEFINITION

BB863087

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 472)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinaawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Toyô, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

TITLE

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 23, 2002, 21:00:44 ; Search time 4203.75 Seconds
(without alignments)
80.267 Million cell updates/sec

Title: US-09-696-686-45
Perfect score: 25
Sequence: 1 agctcagacatgactccatgccc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_estci:*
10: gb_est2:*
11: gb_hc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25	100.0	260	9	BB666889
C 2	25	100.0	406	9	AA002532
C 3	25	100.0	472	9	BB863087
C 4	25	100.0	483	9	BB857917
C 5	25	100.0	485	9	BB857913
C 6	25	100.0	502	9	BB853356
C 7	25	100.0	599	10	BF780486
C 8	25	100.0	608	9	BB605812
C 9	25	100.0	623	9	BB661026
C 10	25	100.0	625	9	BB633565
C 11	25	100.0	637	9	BB650071
C 12	25	100.0	641	9	BB637089
C 13	25	100.0	643	9	BB661603
C 14	25	100.0	649	9	BB652011
C 15	25	100.0	651	10	BE289316
C 16	25	100.0	776	10	BB864846
C 17	25	100.0	779	10	BI688404

C 18	25	100.0	793	10	BI104487
C 19	25	100.0	821	10	BI653380
C 20	25	100.0	838	10	BF534535
C 21	25	100.0	1553	11	AK005532
C 22	25	100.0	1586	11	AK010702
C 23	23.4	93.6	177	9	BB566317
C 24	23.4	93.6	223	9	BB578528
C 25	23.4	93.6	304	9	BB601182
C 26	23.4	93.6	382	9	AA048794
C 27	23.4	93.6	725	10	BG296771
C 28	22	88.0	495	10	BG062828
C 29	20.2	80.8	220	9	BB591610
C 30	20.2	80.8	330	9	BB856691
C 31	19.8	79.2	775	10	BG935945
C 32	19.2	76.8	253	9	AW889975
C 33	19.2	76.8	374	10	BF896723
C 34	19.2	76.8	408	10	BG407897
C 35	19.2	76.8	416	9	AW245217
C 36	19.2	76.8	419	9	AW406598
C 37	19.2	76.8	443	10	BG234205
C 38	19.2	76.8	448	9	AW630286
C 39	19.2	76.8	449	10	BE305036
C 40	19.2	76.8	471	9	AI878843
C 41	19.2	76.8	482	9	AW629814
C 42	19.2	76.8	487	9	AW160553
C 43	19.2	76.8	510	10	BG515497
C 44	19.2	76.8	515	10	BE393833
C 45	19.2	76.8	530	9	AI690580

ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB666889 260 bp mRNA linear EST 26-OCT-2001
BB666889 RIKEN full-length enriched, 2 days pregnant adult female
ovary Mus musculus cdna clone E330035012 5', mRNA sequence.

BB666889

house mouse

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 260)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,

Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda

, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki

, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,

Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,

Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura

, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.